Evaluation, Reasoning and Phenomenal Concepts

Vereker, Alexandra

Awarding institution:
King's College London

Evaluation, Reasoning and Phenomenal Concepts

by

Alexandra Vereker

Dissertation submitted at

King's College London

for the degree of

Doctor of Philosophy

in Philosophy

Abstract

I defend a non-traditional version of sentimentalism about normative reasons for action. I agree with traditional Humeans, such as Blackburn and Schroeder, that desires, or, more broadly, sentiments, are necessary for normative reasons. However, instead of providing a traditional explanation for this necessity (i.e. instead of saying that I have a good reason to do something only if it promotes some desire of mine), I argue that sentiments are necessary for mastering evaluative concepts, and these concepts, in turn, are necessary for having (access to) normative reasons.

In Chapter 1, I show that reasoning alone, understood as coherence and consistency, cannot help us in discovering what we have a good reason to do. There are at least two equally consistent courses of action for a given choice, so one should be equally motivated to do them. What is missing is evaluation, but reasoning alone fails to provide it.

In Chapter 2, I argue, following Quinn and Scanlon, that Humeanism has a problem with normativity: intuitively, it creates reasons where there are none. Just because I want to do something silly, it does not make it any less silly. I argue that to overcome the problem one should admit that desires don't create normative reasons directly, but via providing mastery of evaluative concepts, which then figure in our evaluations.

In Chapter 3, I look at empirical evidence, such as psychopathy and damage to ventromedial prefrontal sector of the brain. Patients with these conditions exhibit emotional deficiencies as well as practical irrationality. I conclude that the best explanation of some empirical evidence is the postulation of a link between sentiments and evaluations.

In Chapter 4, I demonstrate that evaluative concepts are a species of phenomenal concept. Someone who has never experienced colours lacks mastery of colour concepts; similarly, someone who has never had sentiments lacks mastery of evaluative concepts. I argue that this lack of mastery of evaluative concepts is important because such a person would fail to have normative reasons or access to them.

Table of Contents

Acknowledgements

I would like to thank my supervisors Prof. Thomas Pink and Prof. David Papineau, as well as Dr John Callanan, Dr Matteo Mameli and Nigel Vereker for their help. Other philosophical debts are mentioned in the footnotes.

This project was made possible by the funding provided by the Arts and Humanities Research Council and the support of everyone in the Philosophy Department at King's College London.

Introduction

We do things, and often for good reasons. What sort of creatures do we have to be in order for this to happen? This is the question I aim to answer in what follows. The are several traditional answers to this question. The main division between them is whether they emphasize the rational or the emotional side of agents. Various kinds of Kantianism exemplify the first type of answer, whilst sentimentalist theories, such as Humeanism (Blackburn 1998, Schroeder 2007) and sensibility theories (McDowell 1978 and 1979) exemplify the second. In order to illustrate the difference, I introduce a perfectly rational alien who possesses no emotions as all, and call him Mr Spock, borrowing a name of a similar character from the science fiction series Star Trek. According to Kantians, Mr Spock is an agent: he can act, and do so for a good reason. Not so for sentimentalists: they say that an emotional faculty is also needed in order to act for good reasons. The theory I propose is a sentimentalist one: in order to be an agent, one must be both a rational and an emotional creature. In order to act for good reasons, I must be able to evaluate actions. These evaluations contain evaluative concepts. In this thesis, I argue that it is impossible to master these concepts without experiencing sentiments.

I start, in Chapter 1, by showing that reasoning alone fails to provide evaluations. Theories that emphasize rational capacities often have a problem with motivation. I argue that the problem goes deeper – in fact, a purely formal rationalism (one that emphasizes consistency) also faces a normative problem: it cannot explain why one consistent course of action should be preferred over another, equally consistent one.

My theory – I shall call it 'indirect sentimentalism' – is different from sentimentalist theories currently on offer, as I explain in Chapter 2. There, I acknowledge the force of criticisms levelled at sentimentalists, and show how my theory accommodates them. Traditional sentimentalists have a problem in explaining normativity: they move from 'I want to do *x*' to 'I should do *x*'; they hold that I have a good reason to do whatever promotes my desires. But, as their opponents have emphasized, just wanting to do something gives me no good reason to do it. I deal with this problem by denying that

sentiments give one normative reasons directly. Rather, sentiments enable me to master evaluative concepts. These concepts then figure in my evaluations, and evaluations either constitute my normative reasons (if values are not real) or represent them (if values are real). So, my normative reasons are tied to my evaluations, not directly to sentiments. I then spell out the role that sentiments play in mastery of evaluative concepts, in part by making an analogy with phenomenal concepts of colour experience. Someone who is colour-blind, for example, does not have mastery of colour concepts. Her colour judgements may be extensionally correct (she may be able to say what colour objects are if she were to measure the wavelengths of reflected light, for example), yet, intuitively, she is missing something. Similarly, someone who has never had sentiments lacks mastery of evaluative concepts, even though her evaluations may be extensionally correct. The thesis that sentiments are necessary for evaluations is supported by empirical evidence, as I show in Chapter 3. This chapter is devoted to discussion of cases when one's sentiments are abnormal, which leads to abnormalities in evaluations and in patterns of intentional action. I argue that the best way to explain the pattern of action found in some of these conditions is to accept a sentimentalist theory.

In Chapter 4, I argue for my thesis – that sentiments are necessary for mastery of evaluative concepts – by drawing on the work done in philosophy of mind. I show that a creature like Mr Spock, a rational alien who lacks emotions, might have extensionally correct evaluations, yet lacks mastery of evaluative concepts. This means that Spock cannot respond to normative reasons. I have two arguments for this claim. The first argument relies on conclusions reached in Chapter 1: reasoning alone fails to provide evaluations. The second argument spells out the link between sentiments and evaluations. I shall not argue for this link directly, rather, my tactic will be to remove reasons for disbelieving that such a link exists. I show that traditional reservations one may have against sentiment-based theories do not apply to my weakly sentimentalist account. Hence, we have no reason to resist the idea that sentiments are necessary for mastery of evaluative concepts.

I am not taking a metaphysical stance in this thesis: good reasons, or values, for all I

know, may be objective or subjective, or they may not exist. My aim here is to show that sentiments are necessary for agency, whatever metaphysics of values one cares to adopt. They are necessary for agency because our ability to have normative reasons (if values are not real) or to respond to them (if values are real) depends on sentiments.

To recapitulate, my aim in this thesis is to say what agents must be like, and to explain why. In answering this question, I argue against a formal Kantian theory, which says that agents do not need sentiments, and point out a mistake in the traditional explanation of why sentiments are needed. I show that sentiments are necessary for (access to) normative reasons for action not because of the promotion relation, but because of conceptual mastery.

Chapter 1. Rationalism

Part I. Why Reasoning Alone Can't Make Us Act for Good Reasons

1. Introduction

It is often said that rationalist theories of action, such as Kantianism, whilst explaining normativity, cannot explain motivation. In this chapter, I argue that Kantianism has a problem explaining normativity as well, and it is this failure that gives rise to the problem with motivation. This is the task of the first part of the chapter. In the second part, I explain why Kantianism is appealing, and argue that its opponents can retain this appeal. There are many forms of Kantianism, and here I consider only its formal version, as exemplified by O'Neil (1985). I also discuss Smith's (1994) theory, which, one may say, is not purely Kantian, but rationalist.

In the first part of the chapter, I argue that reasoning alone fails to provide evaluations, and hence fails to motivate. This is because reasoning alone tells us only what is consistent and what is not, which is insufficient for motivation. First, I explain how reasoning alone can be thought to motivate: you are motivated to do something when your proposed course of action conforms to the Categorical Imperative, which works by detecting inconsistencies. If there is no inconsistency, then, according to Kantianism, you have a good reason to act as proposed (section 2). Then I pose a problem for the Kantian view by showing that reasoning fails to provide evaluations, and thus leaves us without a definite course of action (subsection 3.1). I then consider two ways in which a Kantian may try to show that reasoning provides evaluations, and argue that they are inadequate (subsections 3.2 and 3.3).

2. A Kantian theory

2.1. Reason and reasons: a disambiguation

There is an unfortunate homonymy in English. 'Reason' can refer to either the process of reasoning or to a reason to do to something, as in 'He left because it was late'. Hume's contention that reason is the slave of the passions is about the process of reasoning (1738-1740, 2.3.3, pp. 413-418). When he argues for this claim, he takes reason to be 'that which judges truth and falsehood' (*Ibid.,* p. 417), a faculty that discovers relations between ideas and causal connections between objects. How does this faculty relate to what we have reasons to do? Let me illustrate. I like chocolate and I think that is a good reason to get it. Here my liking provides me with a reason for action which is most definitely a Humean one. A Kantian would agree so far. But she would add that there is another type of reasons to do things, reached through the process of reasoning alone. If your proposed course of action conforms to the the rules of reasoning (for example, to the Categorical Imperative), then, according to Kantianism, you have a good reason to act as proposed. The debate between Kantians and Humeans is defined by the answer to the question: can reasoning alone provide reasons for action? Kantians answer yes, Humeans – no. A Humean contends that reasoning alone does not give one reasons for action.

There is a wide dissatisfaction with Humean, or, more broadly, sentimentalist theories. The main reason for this dissatisfaction is that our sentiments are contingent – they could have been other than what they are – which makes morality contingent. If responding to good reasons is impossible without having sentiments, and if morality gives us reasons to behave morally, then someone who lacks the requisite sentiment has no reason to act morally. This worries the opponents of sentimentalism: surely, they say, everyone has a reason to be moral, and cruel and insensitive people are no exception. There are also worries connected specifically to Humeanism: in putting motivation first, it fails to explain why I have a *good* reason to do something. Problems for sentimentalism will be discussed in detail in the next chapter, but for now I register them as a *prima facie* reason to seek an alternative. The alternative is a rationalist one.

Rationalists (e.g. Kant 1785 and 1788, Korsgaard 1996, O'Neill 1985) claim that sentiments are not the only source of motivation; reasoning is capable of motivating without help.

So, how can reasoning alone motivate?[1] It motivates, Kant says, when we find that our proposed plan of action obeys the Categorical Imperative (hereafter CI) *'Act only on that maxim through which you can at the same time will that it should become a universal law.'* (1785, 4:421, italics in original).[2] For example:[3]

(1.1) I intend to make a false promise.
(1.2) Universalize this: everyone makes a false promise.
(1.3) If (1.2), then no one would believe a promise, i.e. the practice of making promises ceases to exist.
(1.4) If (1.3), then I can't make a promise.
(1.5) If (1.4), then, *a fortiori*, I can't make a false promise.

The proposed course of action – make a false promise – is not consistently universalizable. The opposite course – always make true promises – is universalizable. Being universalizable is a feature of a good reason, so, if I find that my proposed action is consistent when universalized, I, as someone who cares about acting for good reasons, will have motivation to act on it (Kant 1785, 4:390.) This is how reasoning provides motivation.

Two notes about the CI are in order. First, Kant is not a consequentialist. It may seem that Kant is saying that I cannot rationally want the practice of promising to stop because of the bad consequences this will bring. This is what Singer (1961, pp. 261-275) takes Kant to mean. I think that's wrong. The CI test works on rational

[1] I am very grateful to Dr John Callanan, Michael Campbell, and to the audience at King's College London Advanced Research Seminar for the comments that lead me to understand Kantian theory better.

[2] A disclaimer: it is not my concern to reconstruct the views that Kant himself has held. Rather, I am interested in a theory that claims motivational power to reasoning alone, as defended, for example, by O'Neill (1985) and Smith (1994).

[3] My discussion of the CI follows O'Neill (1985).

consistency, without taking the consequences into account. The point is not that I, or most people, won't want a world without promises. The point rather is that my proposed action is inconsistent when universalized: in a world where false promising is universal, the institution of promising does not exist, so I can't make a promise. Secondly, we must be clear what the CI is intended to preclude. It does not rule out the possibility of acting on motivating reasons: if you made a false promise, you would have acted for a (motivating) reason. The CI is meant to preclude you from doing what you have no good reason to do. If you don't care about acting for good, i.e. normative, reasons, go ahead and make a false promise; but if you do, you should not violate the CI, because, according to Kant, it tells us what good reasons are like – they must be universalizable.

At this point some may worry that the CI test does not get off the ground. It seems that we need sentiments in order to make plans. If we lack them, we lack a plan which can be submitted to the CI test. Reasoning needs the material it can work with: it is only once I *want* to do such and such that I can apply the CI to the proposed plan of action. We can make this worry more explicit with an example. Mr Spock is a perfectly rational alien who lacks sentiments. If the CI applies to all rational beings, it should apply to Spock. But what would Spock do? Will he recognize any possibilities as possibilities for action? Will he make a plan? A Humean says he won't, and it is tempting to agree. A purely rational agent initially seems like a Kantian ideal, but she may fail to be an agent at all. To use Blackburn's snappy characterisation '[w]ithout emotions the will is rudderless' (Blackburn 1998, p. 131).

This worry is easily allayed. A Kantian theory is a theory of normative reasons only. We can leave initial motivation to sentiments, but if we want to know whether we have a good reason to act as motivated, we apply the CI. O'Neill is explicit about this, and, perhaps surprisingly, she agrees with Blackburn that the will is rudderless without emotions and desires:

> The categorical imperative provides a way of testing the moral acceptability of what we propose to do. It does not aim to generate plans of action for those who have none. (O'Neill 1985, p. 259.)

It seemed that the CI applied to any agent simply in virtue of her rationality, so it would apply to Spock. But a Kantian may insist that the CI procedure, although it does apply in virtue of rationality, was not designed with Spock in mind. It was meant to provide a guide for agents like us. Sometimes we fail to act for good reasons because our desires lead us astray. At other times we fail to act for good reasons because we are imperfectly rational, as when we fail to remember some relevant fact and include it in our deliberation. Spock, if an agent at all, will not be led astray or fall victim to irrationality. His agency will be utterly unlike ours, reminding rather that of a holy will (Kant 1785, 4:414). For such a will the CI is not a demand. Its possessor does not have to go through the CI test to find out whether it has a good reason to do such and such; being perfectly rational, it would already know. The CI does not apply to Spock because Spock does not need it, but it is helpful to creatures like us. I am an agent, i.e. someone who cares to act for good reasons. Suppose I have a plan of action prompted by some Humean considerations. I then universalize it. If it can be universalized consistently, then I have a good reason to do it. The recognition that I have a good (i.e. consistently universalizable) reason motivates me:

> for if any action is to be morally good, it is not enough that it should *conform* to the moral law – it must also be done for the sake of the moral law … .(Kant 1785, 4:390, italics in original.)

We have seen how reasoning is meant to provide motivation. As an agent, I care to act for good reasons, the CI tells me which reasons are good, so I am motivated to act on those. In the next section, I pose a problem for this idea.

3. Why reasoning alone can't make us act for good reasons

3.1. Inconsistency vs. evaluation

Blackburn's argument

As we have seen above, the CI is meant to tell me which reasons are good reasons, and

it works by detecting inconsistencies. But detecting inconsistencies is not the same as providing evaluation (and hence motivation) that is necessary for action. It is a fairly uncontroversial assumption that if I act rationally, I have evaluated a course of action. Suppose I believe that education is a good thing. It is this evaluative belief that gives me a normative reason to pursue a degree.[4] This is a natural description, but one that is unavailable if we think of normative reasons in terms of universalization. Some plans of action are consistent when universalized, but what about the evaluation of such plans? Let me illustrate by an example from Blackburn (1998, pp. 217-220). Suppose I intend to pay my credit card off every month. If everyone did that, then banks will not offer credit cards (they are only offered because some people don't pay them off, so banks charge them and make a profit). If there were no credit cards, I could not intend to pay mine off. My intention to pay off the credit card every month results in inconsistency, and hence, according to the CI test, I have no good reason to do this. In fact, since not paying off credit cards on time can be universalized consistently, I have a good reason to do that. But, intuitively, there is a very good reason to pay off credit cards: one avoids debt.

One may object that the examples are different.[5] If false promising is universalized, we can't make sense of the concept of promising at all, because it is essential to promising that I intend to do as I say (at least at the time of the promise). When I say 'I promise to be at the party.', I don't merely raise the probability of going. I am communicating my intention to be there. If I am not doing that, then I don't really understand what promising is. Someone who makes a false promise fails to have the intention, and, since promising involves intending to do as promised, is inconsistent. Compare this with credit cards. If paying off credit cards is universalized, we can still make sense of this concept, because it is not essential that credit cards are offered for profit. All we

[4] For reasons discussed in Velleman (1992), the evaluation that motivates me may not always be a positive one. Velleman denies that agents are always motivated by the good. His counter-examples are Satan, who is motivated by the bad, and depressives, who recognize the good, but are not motivated by it. This is not a problem for my proposal that reasoning alone fails to provide evaluations. Even if we accept what Velleman says, there is still a contrast between evaluations and inconsistencies that my argument rests on. My claim is that any evaluation, positive or negative, is not provided by reasoning alone. Satan may be motivated by the bad, as long as he does not work out which things are bad just through the process of reasoning. Depressives may fail to be motivated, because evaluations are necessary, but not sufficient for motivation.

[5] This objection is due to Dr John Callanan.

understand by 'credit card' is a card which allows you to spend the money you don't yet have in your account. A charity, say, can offer credit cards and not charge people if they fall behind in their repayments. There is no inconsistency in that. The only inconsistency we can get with credit cards, it seems, will be having a card that does not enable you to spend money you don't yet possess, since what we mean by 'credit card' is the card that gives you precisely this ability.

Blackburn responds that when we promise something we don't mean that we intend to keep the promise at any cost. We only mean that we won't change our mind for a trivial reason. For example, it is acceptable to break a promise of buying some bread if my car broke down, and the nearest bakery is miles away (1998, pp. 219-220). Some Kantians agree it may be acceptable to break a promise.[6] For example, we tolerate breaking a promise of going to the party in order to look after a sick friend. This shows that it may be consistent to *break* a promise.[7] But it does not show that it is consistent to *intend to make a false promise*. In making a promise I presuppose the intention of keeping it. This is just what 'promising' means. So, intending to keep the promise is essential to our concept of promising, whilst being offered for profit is not essential to our concept of credit cards. Thus, Blackburn's example fails to show that bad reasons are universalizable, because it does not do justice to what we ordinarily mean.[8]

[6] Kant himself disagrees (1797). He thinks that there are 'perfect duties' – duties that can never be violated – and keeping promises and telling the truth are among those. If a murderer comes to your door asking whether his intended victim is in, and the victim happens to be at your house, you should tell the truth. This, indeed, is an implication of Kant's theory. It is natural to say that we are permitted to lie in exceptional circumstances. But being in exceptional circumstances *does not remove the inconsistency,* which means, according to Kantianism, that one still does not have a good reason to lie. So we could use Kant's own example, instead of Blackburn's, to show that the CI fails to give one intuitively good reasons for action.

[7] There is a problem here: how do we distinguish between the circumstances which let me off fulfilling the promise and the ones which don't? Kantians might say that it is all contained in our concept, but I'm not sure. Some reasons, like looking after a sick friend, are clearly overriding. Some, like simply not feeling like it, are not. But there are plenty of cases in-between, and in some of them our intuitions may be hazy. Thus, Kantians may have a problem in distinguishing between the circumstances which let me off and the ones that don't. But maybe we can sharpen our concepts so that there are no vague cases, so this problem is not as serious as the one discussed in the main text below. The latter problem arises for any concept, vague or not.

[8] As Prof. Pink pointed out, this objection to Blackburn's example is off the mark. Our concept of promising does not require that I have the intention of keeping my promise, but only that I represent myself as having such an intention. I shall, however, continue with an example different from Blackburn's – the example of our concept of marriage – as it allows for easier construction of plausible counter-examples, and does not bring in extra philosophical issues about how promising creates obligations.

My argument

Blackburn's example makes us appreciate an important point: whilst detecting inconsistencies, reasoning alone fails to provide evaluations. We could agree it is essential to the concept of promising that I intend to do as promised. But we can ask: is it the concept we *should* have? After we worked out what our concepts are, we still have two options: either we can keep the concepts as they are, and reject the proposed course of action as inconsistent, or we can change our concepts and go ahead as planned. Reasoning alone is silent about which of the alternatives we should take: it does not provide a motivation to do one over the other. And so it fails to tell us what to do, fails to tell us what a *good* course of action is: it tells us that as long as both courses are equally consistent, they are both equally good.

Let me illustrate with two examples. Suppose we run the CI test on the concept of marriage.

> (1.6) Being already married, I want to marry someone else as well.
> (1.7) Universalize this: everyone who is already married gets married again.
> (1.8) If (1.7), then the practice of getting married ceases to exist. (This is because we pledge exclusivity when we marry.)
> (1.9) If (1.8), then I can't get married.
> (1.10) If (1.9), then, *a fortiori*, I can't get married to someone else as well.

Suppose we run some more CI tests, and suppose they reveal that our concept of marriage is such that marriage is only possible between one man and one woman at a time. This is the concept we have. So, I can either accept the current concept of marriage, and not get married to two people at once. Or I can decide that our current concept is too restrictive, and hence I can get married to one more person. Of course, I cannot change what we mean by 'marriage' single-handedly. But I can campaign for change, and marry two people as an example for others to follow. Whether I succeed in changing the concept is an empirical question, but trying to do so is not precluded by reasoning. Whatever the CI tests reveal about our concepts, we have a choice: we can

either accept the concepts as revealed by the tests, or we can decide to change the concepts. Reasoning alone does not tell us which we should do. Blackburn's response was vulnerable to the objection: this is not what we ordinarily mean by 'promise' or 'marriage'. Mine is not. I agree that the CI may help us understand what we mean by, say, 'promising'. But it does not tell us whether we should change what we mean, and whether this change will make a better course of action available. This gets us to Blackburn's point – that the CI does not tell us what good reasons are. But it avoids the objection that anti-Kantians fail to appreciate what we ordinarily mean by 'promising'.

One may object that reasoning does more than detect inconsistencies, since there are some clear cases when we are motivated to do something after deliberation. So let's go through another example, which does not specifically rely on the CI, but resembles a process of ordinary moral reasoning (Smith 2004, pp. 269-270). Ann wants to give an equal amount of money to Bill and Bob, but less to Charlie. Ann can ask herself why she wants this. On reflection, she finds that she wants to give *x* amount of money to Bob and *x* amount of money to Bill because they are in desperate need, and one should help such people. And she thinks that Charlie is in desperate need as well. So, her belief that people in desperate need should get *x* is inconsistent with her belief that Charlie should get *x* minus *y*. In other words, she realizes she has no justification for treating Charlie differently. In order to avoid inconsistency, she changes her belief that Charlie should get *x* minus *y*, and she acts on it by giving all three people concerned an equal amount. This seems to be a clear case when reasoning leads to a definite course of action.

This, however, is not the *only* reasonable course of action Ann may take. Ann notices that her beliefs are inconsistent, but there are two ways to correct this. Ann can either reject the belief that Charlie should get less than the others. Or she can reject the belief that we should help those in desperate need:

Consistency I
Specific belief: Ann believes that Charlie, Bob and Bill are in need, and that she should give an equal amount of money to each of them.

General belief: Ann believes that she should help people in need.

Consistency II
Specific belief: Ann believes that Charlie, Bob and Bill are in need, and that she should give no money to any of them.
General belief: Ann believes that people should fend for themselves.

A rationalist must show that reaching consistency in the first way is more reasonable.[9] Why should we privilege the general belief over the specific ones? Maybe because of explanatory priority: the general belief explains the specific ones, so Ann should retain it, and specific ones should change to accommodate the explanation. If there is a conflict, Ann should get rid of her belief that Charlie should get less, not of a general belief that anyone in need should be helped. This answer faces the same problem as a second-order desire theory faces. The proponent of a second-order desire theory, Frankfurt (1971), latches onto the fact that we sometimes don't want to desire what we happen to desire. For example, an unwilling addict wants the drug, but also wants not to have this desire: he has a second-order desire about the first-order one. So, Frankfurt tells us, second-order desires are the sensible ones, nothing less than our values. The problem with this is that second-order desires are not qualitatively different from first-order ones: there is nothing preventing one from having sensible first-order desires and paranoid second-order ones. So, in our case, Ann's first-order beliefs can be moral, but her second-order belief can be vicious. For example, Ann may start with a belief that she won't help either Bill or Bob because she has a general belief that people, even the ones in desperate need, should fend for themselves. In spite of this, she finds herself wanting to help Charlie. Being a reasonable person that she is, she realizes that her wanting to help him is inconsistent with her more general belief that people should fend for themselves. So, she still reaches consistency in the second way.

[9] Smith accepts elsewhere that we have two options (2004, pp. 313-314). He says that one is rationally required *either* to get a desire in line with one's belief *or* to abandon the belief. Continuing with our example, Ann has two rationally permissible options: she must either acquire a desire to give the same amount to Charlie as she give to Bill and Bob, or she must abandon the belief that she should help people in need. This does not help the rationalist – surely, we want to say, Ann must go for the first option.

One may object that Ann must have some other reasons not to reach consistency in the second way. She should not have the nasty belief that everyone should fend for themselves. I agree that she should not; but my opponent must say more than 'it is bad that Ann has this belief'. A rationalist must say that Ann can get rid of this nasty belief through reasoning alone. Here is another way Kantians can try to do this. They may agree that second-order beliefs don't help here, but higher order beliefs might. If Ann gets to a general enough belief that any agent must accept on pain of irrationality, she will find a reason not to reach consistency in the second way. She will see that rationality requires her to help others. That is, we are back with the CI. Let's see how an appeal to it can help. The case of helping others as obligatory is spelt out by Onora O'Neill (1985, pp. 275-278):

> (1.11) As an agent, I am committed to the possibility of action.
> (1.12) Some of my actions require help of others as a means of fulfilling them.
> (1.13) In willing those actions I must will to be helped (I must intend the means to my end).
> (1.14) If I universalize the non-helping maxim, everyone (including myself) is not helped.
> (1.15) So, I will both that I am helped (1.13) and I am not (1.14), which is contradictory.

If this argument works, then Ann should get rid of her belief that no one should be helped. She should acquire a belief to the contrary, and provide help where possible.
But the argument does not work. Whilst (1.11) is conceptually true (an agent must be committed to the possibility of some action), (1.12) refers to a *subset* of possible actions: namely, the actions that require that I am helped. But we do not have to commit to the possibility of those actions. O'Neill admits as much:

> It is not a fundamental requirement of practical reason that there should be means available to whatever projects agents adopt, but only that they should not have ruled out all action. (O'Neill 1985, p. 276.)

So, nothing prevents me from ruling out actions which involve help from others: as long as there are some actions leftover, I am not inconsistent. And there would be actions leftover, especially if I count some mental activities as actions – judging, deciding and choosing are likely candidates. Unless we are willing to say that there is no mental agency, the CI does not establish that I am inconsistent in refusing to help others. So it does not prevent Ann from reaching the second way of having consistent attitudes. We find once again that reasoning alone fails to provide evaluations. It only alerts us to inconsistencies, but, like a *reductio* argument, it does not (by itself) tell us which premise is at fault.

Kantians say that the CI identifies good reasons, and we, as rational agents, are motivated by those. But I have argued that because the CI detects inconsistencies, it tells us what our ordinary concepts are. In doing so, it leaves us with an alternative: accept current concepts and follow the CI, or change what we mean and follow the course of action that was inconsistent under old concepts. Reasoning alone does not say which of these we have a good reason to do. So, it has not been shown that the CI motivates on its own. We need evaluations, and reasoning alone does not provide those. Almost everyone agrees that deliberation can *help* us work out what to do. (Hume is an exception. He thinks that passions are not representational states, and reasoning deals only in representations. Hence, reason and passions simply can't talk to each other. (Hume 1739-1740, pp. 415-416.) Neo-Humeans don't take this view.) What is at issue is explaining how reasoning can tell us what to do *on its own*.

So, reasoning alone fails to tell us what to do: for each option, it seems that we have two equally consistent ways of acting, and reasoning alone fails to select between them.[10] We need evaluations, and they are different from requirements of consistency. A rationalist may object at this point that we get this result only because I have adopted an unfairly restrictive conception of rationality: there is more to it than coherence and consistency. This is not a problem for two reasons. First, I agree that the rationalist picture I am attacking is a very formal one. Yet, it is held by authors such as O'Neill (1985) and Smith (1994). They clearly do think that an appeal to consistency can

[10] This paragraph is due to my discussion with members of the audience at the Third Annual Dutch Conference on Practical Philosophy, October 2011, Amsterdam.

explain how we respond to good reasons. Secondly, I think that a certain formalism is a necessary feature of a paradigmatic rationalist theory, i.e. a theory that emphasizes reasoning over emotions. A rationalist theory may be modified to accommodate my objections, but I have difficultly seeing how this can be done without making the theory less rationalistic. I have attacked a fairly formal theory of practical reasoning. Most theories, which I cannot discuss here in detail, are not that extreme, and agree that one's sentiments do play a role in responding to good reasons.[11] I have chosen Kantianism for two reasons. First, because it spells out in considerable detail how one works out whether one have a good reason to do something. Other theories are often less explicit. Secondly, it is unclear how much opposition there is between weak forms of sentimentalism and less extreme rationalist theories. Neo-Aristotelian sensibility theories, like, for example, McDowell's, is one version that a sentimentalist theory can take. So, rather than arguing with friendly theories, I aim to tackle the most extreme, and most obvious, opposition. I correct my pre-occupation with the extremes of the spectrum in the next chapter, where I discuss both a traditional version of Humeanism (Blackburn 1998) and its new, unorthodox development by Schroeder (2007).

However, I have to note that a lot of theories, rather than being pure examples of rationalism or sentimentalism, are hybrids. They are best understood not in terms of *necessity* of rationality or sentiment, but in terms of *emphasis*. I'll illustrate with concrete examples, and I'll use the theories of Damasio (1994) and Smith (1994). Damasio is a sentimentalist: he thinks that emotions, which he takes to be perceptions of bodily changes, are necessary for our ability to mark value and through that, to make rational choices. Michael Smith is a rationalist: he thinks that what we have a good reason to do depends on what agents with maximally coherent, informed and unified desire set would advise us to do. These theories look to be opposing. Damasio emphasizes the visceral changes, whereas rational reflection on, and refinement of, one's desires is important for Smith. However, the two theories can also be seen as complementing each other. The idea that emotions provide information that nothing else provides can be combined with the thought that rational reflection is the way to

[11] Examples of such rationalist theories include Aristotle's theory of a virtuous person's knowledge in *Nicomachean Ethics*, and the medieval model of practical reasoning described in Pink (2004) and (2008).

discover what we have a good reason to do. Smith (and other rationalists can agree with him in this) insists that rational agents must be, amongst other things, maximally informed. So, he could accept sentiments' contribution to normative reasons, because there if we, humans, lacked them, we would not be informed enough. (See also note 23).[12]

So, the debate between rationalists and sentimentalists is less stark than one may think initially. This means that the best way of understanding my thesis is not as having distinctive targets (although there is one – formal rationalism, discussed above), but as showing that sentiments have to be paid attention to, and as explaining why a theory of human agency which does not mention sentiments and their role in deliberation is incomplete.

So, the challenge is: according to formal rationalism, reasoning alone alerts us to inconsistencies, but we need evaluations as well. A Kantian may respond that deliberation does provide evaluations. She may connect what the CI tells us with what is valuable. There are at least two ways to do this.

3.2. Not just inconsistencies – reason discerns values

A Kantian may think that reasoning does not merely detect inconsistencies, it also gives us knowledge of some type of value. This is one way of understanding Kant's argument in the *Critique of Practical Reason* (1788, 5:57-65). He makes a distinction between things that are good and bad depending on the sensations they produce in us (call them sentimental values) and values that are discerned by reasoning alone (call them rational values). If rational values exist, reasoning alone can detect them, and in doing so, it will provide evaluations. In providing evaluations, it will be motivating.

[12] There is a question here about whether rationalists always have humans in mind, or whether they are trying to set conditions for agency *per se*. But, I take it, any rationalist theory, if not specifically about human agency, should at least be applicable to humans, and, as such, would require the modifications which include sentiments.

Why should we accept the distinction between rational and sentimental values? Kant motivates it as follows. Unless we accept values that do not require desires to motivate, we can't have objective values, we can have things that are 'good only in relation to our sensibility' (1788, 5:62). But it makes sense to ask: 'I want X, but is X good?' What we are asking for is a value independent of our desires. If there is such a value, it would be objective and discernible by something other than desires, by reason. So, Kant seems to think that if we accept a Humean theory, then we cannot have objective values. I locate his argument for this claim in the following passage:

> The property of the subject, by virtue of which such experience [of good and evil] could be had, is the feeling of pleasure or displeasure as a receptivity belonging to the inner sense; thus the concept of that which is immediately good would only refer to that with which the sensation of pleasure is immediately associated, and the concept of the absolutely evil would have to be related only to that which directly excites pain.
>
> Even the language is opposed to this, however, since it distinguishes the pleasant from the good and the unpleasant from evil, and demands that good and evil be judged by reason and thus through concepts which alone can be universally communicated, and not be mere sensation which is limited to the individual subjects and their susceptibility. (Kant 1788, 5:58.)

The argument to the conclusion that Humeanism and objective values are incompatible assumes that desires can *only* be contingent. Our desires could have been different from what they are had we been constituted differently. Values can be known either through reasoning or through desires. Given the assumption that desires can only be contingent, values cannot be known through desires because values, if objective, stay the same whatever our desires are and whatever sensibilities we have. Hence, values cannot be discerned by desires. So values, if objective, would be appreciated by reasoning alone.

However, Kantians do not have monopoly on objective values. Although historically sentimentalists (especially Humeans) tended to deny their existence, it is important to point out that they don't have to. I think we should reject the assumption that desires can only be contingent. How we know the world depends on what sort of things the world

contains. If it contains solid objects, they can be known via touch (by those creatures who possess this sense). If it contains values they can be known via desires (by those creatures who possess this faculty). In this case, some desires will be not contingent, but necessitated by our acquaintance with objective values. (Our possession of the faculty of desire is, of course, contingent. But the same can be said about the faculty of reason. Here both faculties are on a par.) Philosophers such as Plato, Oddie (2005) and Mackie (1977) rejected contingency assumption, but Mackie was the only one out of the three who rejected objective values; Oddie and Plato are value realists. For Oddie, desires are data for what is valuable, just as perceptions are data for what is true. Seeing a red rose gives me a *prima facie* reason to believe that it is red. Wanting a bit of cake gives me a *prima facie* reason to believe that it is good.[13]

Plato's name maybe a surprise here. After all, he is famous for his rationalism, so I shall make a short digression to show that he is not a rationalist of the formal kind. This will also help further to explain the idea of desires as responses to objective values. There are good reasons for classifying Plato as a rationalist. He does talk of the body and its desires as weighing us down like an oyster shell (*Phdr.* 250c). His rationalism is also evident in the passages on the tripartite division of the soul (*Phdr.* 246a-c, 253d-256d, *Rep.* 436a-444e). The three parts of the soul are: rational, spirited (which is often taken to correspond to emotions) and appetitive. In an ideal soul, Plato tells us, reason rules, subduing appetites and emotions. But there are a couple of passages that sit ill with this straightforwardly rationalist picture.

The first passage is the erotic ascent in the *Symposium* (*Symp*. 210a-212b). In Greek there are two words that correspond to the English 'love': *eros* and *philia*. *Eros* emphasizes sexual desire, whereas *philia* connotes affection, such as love for one's friends and family (Vlastos 1973, n. 4, p. 4). In the *Symposium* Plato talks exclusively about *eros*, so in what follows I shall replace the broader term 'love' with 'sexual desire'. The ascent (*scala amoris*, in Vlastos' memorable phrase) described in the *Symposium* is

[13] Of course, the claim that desires are necessitated by acquaintance with objective values has an 'and all goes well' clause: favourable conditions are assumed. I may be in the presence of a red rose and not see it because my sight is defective or because the rose is somehow obscured. I may be in the presence of value and not desire it because, for example, my desire faculty is malfunctioning or because the value is somehow obscured. (I owe this point to Nate Sharadin.)

from one sexual object to another, and is enabled by the object's possession of beauty. When I see something beautiful, I have *eros* for it. At first, one lusts after one beautiful body, then realizes that the same quality of loveliness is possessed by other bodies and comes to desire them too. Then one becomes aware of the beauty of the mind, and treasures that more than bodily beauty. The next step is erotic desire for beautiful institutions and pieces of knowledge, until at last one sees Beauty Itself, which, unlike the beauty of any particular thing, is eternal, and never partakes of ugliness.

In this surprising to modern ears passage, Beauty Itself is no less an object of sexual desire than a beautiful body. In fact, 'no less' is an understatement. If your sexual desire can be aroused by an imperfect beauty of some perishable physical body, how much more *eros* would you have for Beauty Itself were you to see it! This is a vivid description of objective values and what it would be like to perceive them. If one were to see them, they would affect not only the intellect, but also the appetite. You see Beauty Itself, albeit with the mind's eye, and you *want* it because it is so beautiful. This is so striking that it requires elaboration: does Plato really think that we have sexual desire for Beauty Itself? Well, at the lowest level of the ladder, when the object of sexual desire is a beautiful body, there is no temptation to say that *eros* is reasoning personified. Now, the word – '*eros*' – has not changed throughout the ascent to other objects. This is one reason to think that *eros* is still the same sexual desire applied to a different object – Beauty Itself. Also, sexual desire is a response to the property 'beautiful', and Beauty Itself possesses it to the utmost degree, so sexual desire would, if anything, be stronger in the case of this supremely beautiful thing than in the case of the objects lower down the ladder. The property (beauty) has changed in degree, not in kind, so there is no reason to think that the erotic response has changed to some other kind of response. We may, however, question the idea that the property has not changed in kind: after all, why suppose that when talking about the beauty of intellectual objects we are talking about the same thing as when we are talking about a pretty boy? But it seems that Plato does suppose this. He thinks that desiring bodily and intellectual beauty helps us ascend to Beauty Itself. In order to enable this ascent, the property of beauty must be the same property throughout. Beauty possessed by a boy will get us to the same thing, but more refined. It will not get us to some unearthly beauty which is a

completely different property. If beauty of the boy and beauty of Beauty Itself are different in kind, then the ascent Plato envisaged is impossible.

The second passage that makes one wary of straightforwardly classifying Plato as a rationalist comes from the *Republic*:

> it is in the nature of the real lover of learning [i.e. philosopher] to struggle toward what is, not to remain with any of the many things that are believed to be, that, as he moves on, he neither loses nor lessens his erotic love until he grasps the being of each nature itself with the part of the soul that is fitted to grasp it, because of its kinship with it … (*Rep.* 490b.)

Unfortunately, we are not told which part of the tripartite soul (reason, spirit or appetite) is talked about here, but *eros* for the idea is clear enough. We are also told that the *eros* we have for beautiful things would be nothing in comparison to the desire for wisdom, were it accessible to our sight: an image of wisdom 'would awaken a terribly powerful love' (*Phdr.* 251d).[14]

So, objective values don't have to be appreciated by reason alone. Kantian monopoly on them follows only given an additional assumption: that desires, or sentiments, can only be contingent. I have argued above that if we accept that values are objective, we have no reason to agree with this assumption. Sentimentalism is compatible with objective values. (I do not take a stand on metaphysics here. I merely show the compatibility.)

If sentiments can tell us about objective values, there is no motivation to divide values into rational and sentimental ones. Moreover, there is motivation not to divide values. Sentimental values, in so far as they are values, excite our non-cognitive faculties. This leads to motivation in their case. It is part of our concept of values that they are motivating. It is natural to say that if you are not motivated to get away from dangerous things, you don't know what 'dangerous' is. Typically, if you say that something is the right thing to do, yet are not at all motivated to do it, then you are 'just saying it' without meaning what you say. Obviously, there is some slack: we are not always motivated to

[14] I am not pretending that I found a consistent reading of Plato. But presenting him as a rationalist should not let us ignore the passages where Plato clearly states that a philosopher has *eros* for the idea.

do what we value. (See Stocker (1979) for good examples.) The slack shows that the token claim – that each time I sincerely make a value judgement I must be (to some extent) motivated – is false. So we must retreat to the level of types – when I make a sincere evaluative judgement, this type of judgement is distinguished from other types by the fact that it will, when things go well, motivate me. We can easily understand someone who *on this particular occasion* is not motivated by her value judgement, but not someone who claims to hold a certain value, yet is *never* motivated to act in accordance with it. Of this last person we would say that she does not really value what she claims to value. So, in so far as rational values are values, they will be motivating. Why think that in their case motivation is provided by something different than in the case of sentimental values? To rephrase: we know that values motivate, and motivation in the case of at least some familiar values is provided by desires. If there is some other stuff that belongs to the *same* category (values, motivating things), we have a defeasible reason to treat it in the same way. But even a defeasible reason is a good reason, unless a contrary one is provided. The example of a rational value that Kant has in mind is moral value. As he points out, it makes sense to ask 'X is pleasurable, but is X good?' (1788, 5:58). But it also makes sense to ask 'X is rational, but is X morally good?'

Another reason why Kantians may think that desires can't provide information about values is that they see desires as non-representational passions.[15] In the Groundwork (4:399), Kant distinguishes between practical and pathological love. Practical love resides 'in the will and not in the propensities of feeling', and can be commanded. Pathological love is a type of feeling, and cannot be commanded. Presumably, it cannot be commanded *because* it is a type of feeling, or sensation: it cannot be commanded because it has no representational qualities, and hence cannot respond to reasoning and commands. Perhaps surprisingly, this is how Hume thought of desires, as well. When I am angry, he says, I am 'possessed with the passion, and in that emotion have no more reference to any other object, [than] when I am thirsty, or sick, or more than five foot high' (1738-1740, 2.3.3, p. 415). But whatever stand one takes in the Kantian/Humean debate, there are independent reasons to reject the claim that passions are non-representational. To continue with Hume's own examples, it looks like sentiments are

[15] I thank Prof. Pink for this point.

object-directed: when I am angry, I am angry at John, or at an injustice, or at the state of the world. When I am thirsty, I want a drink, or a Fanta, or that glass of water. Sickness, on the other hand, is markedly unlike this: it is not directed at any object.[16] There is nothing in the neo-Humean position that forces one to abandon this common-sense view. The same goes for Kantians. They can admit that passions represent, and still retain their opposition to the idea that passions have a role in discovering values, because passions can only be contingent. This line of thought, however, was attacked above: one need not think of passions as contingent should the world contain values.

Before I go on, I shall consider the following objection. I have described desires as if they were perceptions of values, should values be objective. I have also used terms such as 'sentiments' and 'emotions', and assumed that they play the same role in a sentimentalist theory that desires do. However, at least *prima facie* desires and emotions are different from perceptions. The former two are reason-responsive (desires, possibly, more so), whilst the latter are not.[17] This reason-responsiveness is demonstrated by the fact that I can argue you into having a desire to go out tonight by presenting you with various tempting options; the same does not seem to hold for perceptions (and, sometimes, emotions). In the case of emotions, if you are afraid of a spider, I may be unable to argue you out of it: you can continue to be afraid, even though you have a rational belief that the spider is harmless. In the case of perceptions, you may firmly believe that the two lines in the Muller-Lyer illusion are the same length, yet this does not affect how they look to you: the line ending with arrows looks shorter than the line ending with forks. Thus, it looks like desires must be reason-responsive on pain of irrationality, perceptions are not reason-responsive, and emotions are somewhere in the middle – sometimes they are reason-responsive (for example, I can argue you into being angry about a building project by convincing you that it is unfair), and sometimes they persist independently of one's judgements (like in the harmless spider example).

[16] However, there are theories that make all mental states, including sensations, representational. Sensations, the proponents of representationalism argue, represent the state of my body (e.g. Tye 1995).

[17] I owe this objection to Prof. Pink.

In response, I note first, that this distinction is not uncontroversial. If one defines desires by their reason-responsiveness, some of the states commonly known as desires will not qualify as such. One cannot, for example, argue someone out of a desire to drink. The proponent of reason-responsiveness of desires would say that thirst is a mere feeling which is, of course, not reason-responsive. This may be true, but there are other examples. A desire for a cup off coffee (even when I am not a caffeine addict) may have a strong phenomenology and may fail to be reason-responsive. My desire to get a new dress may be defeated, in the sense of not leading to an intention, by reasons such as lack of money, yet it does not dissipate even after I acknowledge the need for frugality. Scanlon, whose work is discussed in more detail in the next chapter, has captured this feature of desires by introducing what he calls 'desires in the directed-attention sense' (Scanlon 1998, p. 39). When I want something, my attention is persistently grabbed by the attractive features of the object of my desire.[18] Such examples are ubiquitous and easily recognizable, so we may conclude that persisting in the face of contrary considerations is no less characteristic of a desire than (certain) reason-responsiveness. This point is reinforced by the observation that both desires and perceptions are perspectival, which I discuss in Chapter 2, section 4.

One could also argue that perception can be reason-responsive in the following loose sense. Noe (2006) argues that it is 'bad phenomenology' to think about our visual experiences as snapshots. Instead, what we see depends on what experiences we'll have when we turn our heads slightly, focus on one element or another, move around the perceived object, and, crucially for my purposes, on the experiences we *think* we shall have if we were to do all these things.[19] If what we perceive depends on what we expect to perceive, then it is a kind of reason-responsiveness. This thesis may seem controversial and implausible. However, I have a confirmation of it from my own

[18] Scanlon, of course, thinks that these attractive features are reasons, in which case desires are responsive at least to *pro tanto*, if not always to *pro toto* reasons. But one could accept that these features grab our attention independently of our judgement about what is best, yet deny that they are perceptions of reasons, as Scanlon claims them to be. For example, my attention can be grabbed by these attractive features not because they are reasons, but because I have conflicting desires: a taste for expensive clothes and a desire not to go on a spending spree quite so often.

[19] Noe's aims are different from mine. He does not say that perceptual experiences are reason-responsive. Rather, he argues against internalism about mental content, i.e. the thesis that what mental states I have depends only on my intrinsic properties. Noe's (externalist) view is that in order to have some mental states one has to be 'related to the environment in the right way'. (Lau and Deutch 2010.)

phenomenology. When staying in the Lake District, I admired the view from our window. The two hills that I could see looked about the same height and the same distance from Lake Windermere. They looked so nice that we decided to walk up there, and so we did. On coming back after the walk, I looked out of the window again, and was extremely surprised: the view looked different. Now one the hills looked shorter and closer to the lake, whilst its neighbour was clearly higher and further away, as our walk has proved it to be. The difference was startling. Thus, visual perception can depend on one's expectations of what one will see after she moved in a certain way, and is reason-responsive in this looser sense.

Even though the objection does not succeed, because it is unclear that desires are reason-responsive to an extent that cannot be matched by emotions and perceptions, it leads to a related question. What I designated as 'sentiments' is quite a broad term, so what unifies the items under it? My answer is that the criterion for identification here is phenomenology: in order to count as a 'sentiment', in my sense, a mental item has to have a characteristic feel.[20]

3.3. Not just inconsistencies – reason creates values

A Kantian can make a different response. Reason does not provide access to a special type of value. Instead, it *creates* values. I make the object valuable by choosing it:

> … what makes the object of your rational choice good is that it *is* the object of a rational choice. (Korsgaard 1996, p. 122, italics in original.)

A natural question arises: if I confer value to the object by rationally choosing it, what makes my choice rational? Korsgaard may explain the rationality of my choice by reference to other rational beings (cf. 1996, p. 241). My choice of A

[20] This may exclude some desires. Schueler (1995) claims that some desires, such as a desire to visit my sister, lack phenomenology. This desire is just a disposition to do certain things. I am not sure whether any desire can be reduced to a disposition: just because a desire lacks the strong, distinctive phenomenology of, say, sexual desire, it may still have some phenomenology. (At least it does when I imagine that I want to visit my sister.) But if some desires do lack phenomenology altogether, they are excluded from my definition of sentiments. This is not a problem, because, as I argue in this chapter, a rational creature only armed with consistency and coherence will fail to have even purely dispositional desires. No desires – even purely dispositional ones – are based on reasoning alone.

over B is rational if anyone rational chooses A over B after correct deliberation. This explains why my choice counts as rational – because it follows the laws of rationality, and what I chose is good because I rationally chose it. But this explanation gets things backwards. To show why, I use Plato's discussion of piety (*Euth.* 7a-10d). Socrates asks Euthyphro what pious is. Euthyphro answers that pious is what gods love. Then Socrates asks: 'Is the pious being loved by the gods because it is pious, or is it pious because it is being loved by the gods?', and answers his own question: it is loved by the gods because it is pious, not vice versa. The same goes for rational choice – if I choose rationally, it is because the object of my choice is valuable, not the other way round.

One may object that the comparison is unfair. Gods may happen to love anything, which may or may not turn out to be pious. But laws of rationality will prevent me from getting any odd thing as my rational choice, they will keep me in line. The procedure by which I arrive at my choice – reasoning – is not arbitrary. This objection is countered by what has been said above, where I showed that the process of reasoning alone will not lead us to a particular choice. I may discover that my concepts are inconsistent, but I still have two options: keep the current concepts and follow the course that avoids inconsistency or change the concepts I have and avoid inconsistency that way.

4. Conclusion

Reasoning alone alerts us to inconsistencies in our plans of action. In doing so, it fails to select between courses of action that are equally consistent. We need an additional standard – that of evaluation – in order to select between courses of action that are equally consistent. So, a purely formal rationalist theory has a problem with normativity: it fails to provide a way of deciding between consistent courses of action, even though intuitively one is better than the other.

Part II. Kantian Appeal Explained

1. Introduction

In the first part of the chapter, we have seen that rationalism has a problem: reasoning alone fails to provide evaluations. However, rationalism has been, and continues to be, appealing. And my task will not be complete without explaining its appeal, and showing that a sentimentalist alternative can retain it. My example of a sentimentalist theory will be Humeanism. I argue that Kantianism is appealing because it satisfies our pre-philosophical intuition about normative reasons for action:

> For any normative reason to φ: if some feature F in circumstances C gives me a normative reason to φ, then F will give a normative reason to φ to any agent in C.

I show that Humeans can also accept this feature of normative reasons, thus accommodating the way we ordinarily think about them at least to some extent. I then argue, however, that traditional Humeans fail to put enough space between motivations and normative reasons. This can be shown via disagreement about which ends one should pursue. I introduce my solution to this problem – I use evaluations as a way of distancing the motivational from the normative – which will pave the way for the indirect sentimentalist account spelt out in the next chapter.

2. A Kantian intuition: nothing special about me as user of reasons

2.1. A Kantian intuition

Suppose I am walking in the park. It is a nice, hot day and I rather fancy an ice-cream. I have no reason not to buy an ice cream (I am not allergic to it, I don't need the money to buy bread for my children, etc.). In this case I have a reason, and a good reason, to buy

an ice-cream. It seems that as long as it is a reason, and a good one, if *you* were walking in the park on a nice, hot day, and fancied an ice-cream, and did not have a reason not to buy it, then you would also have a good reason to buy an ice-cream. If my wanting in circumstances C gives me a good reason to φ, then, if this really is a good reason, it will function in the same way for a different agent in the same circumstances. Whether it is you or me who is walking in the park wanting an ice-cream is irrelevant for having a reason. This seems to be a feature that all normative reasons have. I put this intuition more formally:

> *Normative Reason Feature (NRF)*
> For any normative reason to φ: if some feature F in circumstances C gives me a normative reason to φ, then F will give a normative reason to φ to any agent in C.

This is a universal constraint on normative reasons: any normative reason must satisfy it in order to count as one.

Note three things:

1. Humeans accept this constraint, but they make a restriction on the feature that gives us normative reasons: for Humeans, this feature F is always a desire.
2. The NRF itself is not relative to an agent's desires; it applies independently of any desires she may have.
3. This constraint is expressed by a conditional, which can be true even if nothing satisfies it: that is, it can be true even if there are no normative reasons at all.

The NRF still allows for differences in personal taste. For example, suppose Mary likes wine and John hates it. There is a bar nearby that serves a great selection of wines. Intuitively, this is a reason for Mary, but not John, to go to the bar. According to the constraint, Mary and John agree on the following:

> *if* the fact that there is good wine at that bar gives Mary, in her circumstances (where these include liking wine) a normative reason to go to that bar, *then* the fact that there is good wine at the bar would give John a normative reason to go to

the bar if he were in the same circumstances as Mary (that is, if John liked wine as well, which he does not).

So, the constraint does not mean that John has a reason to go to the bar, even though he hates wine. It only says that *if* Mary's liking of wine gives her a reason to go to the bar, *then*, if John liked wine, he would also have a reason to go. In actual fact his circumstances are dissimilar from Mary's (he does not like wine), so he does not have a reason to go to the bar.

Moreover, and this is easily overlooked, the constraint proposed does not say whether, given her liking wine, Mary *actually* has a reason to go to the bar. This is because the constraint is not meant to be a full account of what it is for someone to have a reason. It is just one plausible constraint on normative reasons, there may be others. It is not meant to indicate what reasons everyone has, or even whether there are things that everyone (or anyone) has a reason to do. *It is just a parity principle*. *If* you assume that

> there are normative reasons
>
> and
>
> some feature F – e.g. your liking vanilla ice-cream, to use Sobel's (1999) example – gives you a normative reason to buy this flavour in circumstances C,

then you must, on pain of inconsistency, admit that

> this same feature F – another agent's liking vanilla ice-cream – gives that agent a normative reason to buy this flavour in circumstances C.

The NRF formalizes an intuitive thought: we do think that good reasons are independent of at least some features of the agent. There is nothing special about me as a user of reasons. Cf. beliefs: I can't rationally believe any odd thing I like; if my beliefs are to be rational, they have to satisfy constraints that are independent of me. Similarly for normative reasons: I may have motivations to do all sorts of things, but if they are to be good reasons for action, they have to conform to constraints which are independent of me, the agent, and are the same for all agents. It is this thought that rationalism, and Kantianism in particular, tries to accommodate, and it is this thought that is expressed

by the Normative Reason Feature.

Beliefs	+	Constraints	➔	Rational beliefs
Motivations	+	Constraints	➔	Normative reasons

At this point it may be useful to re-state the question I am trying to answer: what sort of creatures do we have to be in order to do things for good reasons? Thus, I am not interested in moral action in particular; egoism, for example, is consistent with my thesis. There may be some creatures that only respond to reasons connected to their self-interest, but even of such beings we may ask: what has to be true of them? And my response, developed in subsequent chapters, is that such creatures must have sentiments; reasoning alone is not enough to respond to reasons, moral or otherwise.

We have identified the intuition behind Kantianism. At what level of understanding of normative reasons can this intuition be accommodated? Let us consider three agents, each exemplifying a different level of understanding of normative reasons.

Level 1 – Extreme egocentric.
Someone who takes only her own reasons to be reasons. E.g.:

A: Can you lend me some money, I need to buy a present for my daughter?
B: No, I need it to buy a present for my own daughter.
A: Why would I care about your daughter?

In this first case we have someone who thinks that her reasons are much weightier (better *qua* reasons) than anyone else's.

Level 2 – Humean agent (sensible knave).
I have a reason to coerce you into doing what I want, and I appreciate that you have a reason to resist. I also appreciate that you have a reason to coerce me to promote your interest, and I have a reason to resist you then. There is a kind of universalization here: the knave admits that if promoting her interest is a reason for her, it is also a reason for

anyone. This is what I call 'Humean universalization'. In this second case we have someone who thinks that everyone's reasons have equal weight: my reason is as good as yours in so far as they are good reasons.[21] Agents' reasons may clash (A will coerce B to promote her interest, B will coerce A to promote hers), but, since they are of equal weight, there is nothing to decide between them in terms of rationality. Humeans may still say that A and B have a reason not to coerce each other, because it is *bad* to treat someone like this. 'Bad' here does not refer to incorrect reasoning, but to our appraisal based on our values.[22]

Level 3 - Kantian agent.

I have a good reason to do something only if everyone has a reason to do this. So far this is no different from the sensible knave. We need to add: I discover good reasons through reasoning alone. I can do this, for example, by working out whether a world where everyone acts as I propose is consistent. Such discovery of reasons by applying the rules of reasoning, and, in particular, of consistency, is a feature of a formal rationalist theory, as proposed by O'Neill (1985) and Smith (1994). This is what I call 'Kantian universalization'. In this third case, we have someone who thinks that anyone's good reasons have equal weight, and she tests for a good reason by seeing whether a world where everyone acts on this reason is consistent. A and B do have a reason not to coerce each other, for example, because it is *inconsistent* to treat each other like this.

So, we have three cases. In which of these is the NRF satisfied? Level 1 clearly does not satisfy it. Someone who treats only her reasons as normative reasons does not understand what normative reasons are. The agent at Level 1 fails to accept that there is nothing special about her as an individual in relation to reasons. She takes her motivations to be all and only good reasons, and this is not what good reasons are. The agent at Level 2 does satisfy the NRF: whatever gives a good reason to the knave, gives a good reason to anyone. So, if the NRF captures the intuition behind Kantianism, we

[21] Is this equal weight measured from the point of view of the agent, or from the point of view of an impartial observer? I believe that in order to appreciate what reasons are, one has to be able to, at least sometimes, take the all-agents'-reasons-are-equal, impartial point of view. This is why an egocentric comes across as someone who fails to appreciate what good reasons are. However, appreciating that your reasons are as good as mine does not rule out the sensible knave's position, as I argue below.

[22] This is the approach taken by Blackburn (1998).

don't have to go beyond Level 2 and introduce a further constraint – that a good reason must be universalized consistently. Kantians go too far in trying to accommodate ordinary thought.

Kantians, of course, say that the knave is failing to appreciate what good reasons are. Here is how Smith argues for this conclusion (1994, pp. 193-196). For Smith, one has a normative reason to φ in circumstances C iff one's fully rational self (i.e. a self that has a maximally informed, coherent and unified desire set) would advise one to φ in C. (Smith 1994, pp. 151-152, 2004, pp. 263-264.)[23] A knave, or 'successful criminal', as Smith calls him, thinks that he has a normative reason to gain wealth no matter what the cost to others:

> as we have seen, this is equivalent to the claim that fully rational creatures would want that, if they find themselves in the circumstances of the successful criminal, then they gain wealth no matter what the cost to others. And the successful criminal's opinion notwithstanding, it seems quite evident that we have no reason to believe that this is true. Fully rational creatures would want no such thing.
>
> Note what I have not said. I have not said that the fact that we all disagree with the successful criminal entails that he is wrong. Perhaps we are all mistaken about what fully rational creatures would want. But the mere fact that it is logically possible that we are wrong gives us no more reason to endorse the opinions of the successful criminal and doubt our own convictions than the mere fact that it is logically possible that we are wrong when we think that the sun will rise tomorrow gives us reason to endorse the opinions of the prophets of doom. (Smith 1994, p. 195.)

So, the successful criminal (the knave, as I call him, and I'll stick to the masculine pronoun as Smith does) thinks that fully rational creatures in his circumstances would want to gain wealth no matter what the cost to others. But Kantians think that fully

[23] Smith's definition glosses over the differences between sentimentalism and rationalism. When does one have 'maximally informed' desires? In the following chapters I argue that, without sentiments, one would lack mastery of evaluative concepts. If this is true, then Smith could admit that the desires of someone who lacks sentiments fail to be maximally informed. This admission would mean that Smith no longer holds a formal version of rationalism, which emphasizes reasoning over sentiments. If he were to concede a greater role of sentiment in the process of responding to normative reasons, I no longer have a quarrel with him. I do, however, take Smith's theory to be Kantian in spirit, because he thinks that convergence of agents' desires will be achieved via a process of reasoning, understood in terms of consistency and coherence.

rational creatures in the circumstances of the knave would not want that. They appear to be at loggerheads. But, Smith says, we have a reason to prefer Kantian opinion to the knave's. Given everything else we believe about the world, we shouldn't believe the world will end tomorrow. Given everything else we believe about how we should treat each other, we should not believe that it's OK to gain wealth no matter what the cost to others. I agree with Smith that we should not believe this. But in order for this observation to support a Kantian position, we need to show that beliefs about how we should treat each other have a rational, rather than a sentimental, foundation, and this is precisely what is currently at issue.

But this is a side point. Kantians require the knave to be irrational. Smith thinks that the knave is irrational because

> he sticks with his opinion despite the fact that virtually everyone disagrees with him. Moreover, he does so without good reason. For he can give no account of why his opinion should be privileged over the opinion of others; he can give no account of why his opinion should be right, others' opinions should be wrong. (*Ibid.*)

And so, Smith concludes, the knave is intellectually arrogant. He discounts the arguments of others without producing counter-arguments. And whoever does this is irrational.

Indeed, some descriptions of the knave can be filled out to make him irrational, but then the knave is no longer as I characterized him. Suppose the knave does not want others to do to him what he proposes to do to them (Foot 1972, p. 161). This is not enough to decide whether the knave is inconsistent, but the case can be constructed so that he is. Suppose he thinks that others have no normative reason to treat him as he treats them, whereas he does have such a reason, simply in virtue of being himself. Such a person is, indeed, inconsistent, but this person is not the knave, according to my definition. He is failing to appreciate the Normative Reason Feature: he only takes his own reasons to be reasons. This is the extreme egocentric – level 1, not the knave, who is at level 2. The knave accepts the Normative Reason Feature. He realizes that there is nothing special about him as a user of reasons, so if he has a good reason to treat others badly, so do they. And if he has a good reason to resist coercion, so do they. The knave accepts that if any rational agent were in his circumstances, then she would have a

normative reason to treat others badly. In accepting this, he is not behaving as someone who discounts others' opinions. He admits that anyone's normative reasons are as good as his. He can even admit that he may be mistaken about what we have a normative reason to do. But if he does, so should a Kantian. In Smith's terms, if everyone started off as a knave, and then, through correct deliberation, became a moral person, then Kantians are correct. If everyone started off as a moral person and then, through correct deliberation, converged on the knave's desires, then the knave is correct. But at present we don't have such convergence, so we don't know who is correct.[24] For Occamist reasons, we may side with the sensible knave. He only requires a minimal (level 2) convergence in order for normative reasons to exist. The rationalist makes a bolder claim and requires level 3 convergence, and if we don't have that, then there are no normative reasons.

What one should note is that sentimentalism does not preclude convergence in everyone's desires. As we have seen in Part I section 3.2, sentimentalism is compatible with objective values. If there are such things, we may converge on them, but not, according to sentimentalist, by a purely rational process. Instead, such convergence will be, at least partly, a matter of sensitivity. Convergence could be provided by acquaintance with objective values, in which case you don't have to be fully rational, you just have to be fully receptive to values, and that receptivity does not have to be cognitive.

The NRF is prefaced with 'for any normative reason', so it is a universal constraint on good reasons. Therefore, the discussion above leads to an important distinction between universality of constraints on reasons and universality of reasons themselves. Kantians (and not only they) often talk about reasons that everyone has. When I explained the Kantian intuition, I used the phrase 'nothing special about me as a user of reasons'. This may be reminiscent of what is called 'universality of reasons'. I have not used this phrase because I think it's misleading. When we call reasons 'universal', or 'reasons that

[24] Smith (1994, pp. 187-189) is optimistic about the possibility of Kantian convergence because of the level of moral agreement we have reached historically. Sobel (1999, pp. 145-147) offers persuasive, in my view, reasons why much historical moral agreement is not helpful for a rationalist. In brief, agreement has often been reached for the wrong reasons. Arguments were accepted because of the position of those offering them, not because arguments themselves were good. And sometimes arguments were accepted because they were rationalizations offered after the prevailing attitudes have already changed. The best case for rationalism is convergence among societies that differ in the evaluative concepts they use, and Smith has not convinced us that there are any such cases.

everyone has', 'reasons that we all share', we end up with

> Universality: Everyone rational has a reason to … [purely in virtue of their rationality].

The dots can be filled in by different things: promote her own welfare/promote the greatest happiness/respect others. This is very different from the Normative Reason Feature. The NRF does not require us to get above the level of the sensible knave. Universality asserts that there are some reasons that any rational agent has simply in virtue of being a rational agent. It goes to the level of a Kantian agent. The NRF, on the other hand, is a conditional. It does not say, unlike Universality, that there are some *reasons* that everyone shares. It only says that if there are good reasons, they are subject to constraints which are independent of the agent who uses them: there is *a feature* common to all good reasons. Unlike Universality, the Normative Reason Feature says nothing about sharing reasons themselves.

2.2. Disagreement about reasons[25]

In this subsection, I shall discuss another argument which is Kantian is spirit, and show that, although the argument can be countered by Humeans, it still points to a real weakness of traditional sentimentalist theories: they fail to provide enough space between normative reasons and sentiments. I then suggest that one way of providing the necessary space is to introduce evaluations, and make them, and not sentiments themselves, act as normative reasons. I shall explore this option more fully in the following chapter.

Smith (1994) argues that our pre-philosophical conception of normative reasons is anti-Humean. For him, the debate between Humeans and their opponents is about whether or not normative reasons are relative to our desires. The definitive feature of anti-Humeanism is that

> under conditions of full rationality we would all reason ourselves towards the same conclusions as regards what is to be done; … via the process of systematic

[25] I thank audience at the Société de Philosophie Analytique (SOPHA) 2012 conference in Paris and as well as Prof. Pink for their comments on this subsection.

> justification of our desires we could bring it about that we converge in the desires that we have. (Smith 1994, pp. 165-166.)[26]

Humeans oppose this view by holding that

> desires an agent would have if she were fully rational are themselves simply functions from her actual desires … An agent's reasons are thus relative to her actual desires … because we cannot expect that, even under conditions of full rationality, agents would all converge on the same desires about what it is to be done in the various circumstances they might face. (*Ibid*., p. 165.)

So, anti-Humeans à la Smith hold that as long as we deliberate well, we would all end up with the same desires; we converge, and it matters not what desires we start with. Their conception of a normative reason is non-relative. Humeans, on the contrary, say that there is no guarantee that we would all converge on the same desires, however well we deliberate; what desires we start with does matter. Their conception of a normative reason is relative. (For Humeans, not only actual desires, but even hypothetical ones may fail to converge. To use Williams' (1979) famous example, someone who wants to drink the contents of a glass thinking it is a gin and tonic would not want it if he knew that the glass contains petrol. Her desire to drink from this glass would dissipate had she known the relevant facts. However, Humeans hold that such change in our desires in no way guarantees that we will all end up with the same desires had we taken account of all the relevant facts.)

Smith then argues that if we accept the Humean conception, then we can't disagree about reasons.[27] When I (a Humean) say that you don't have a reason to go on holiday, I am talking about reasons-relative-to-*my*-desires, and when you (a Humean) say you have a good reason to go on holiday, you are talking about reasons-relative-to-*your*-desires. Thus, we are not disagreeing, we are talking about different things. According to a non-relative conception of reasons, we are talking about reasons *simpliciter*, so we can disagree:

---

[26] As Sobel (1999, p. 139) points out, Smith requires *de se* convergence, i.e. a situation where we both want the biggest piece of cake for ourselves, not a situation when I want the biggest piece and you also want me to have it. So, Smith's convergence does not preclude the possibility of egoism: it may be that all agents converge on a desire to promote their own interest.

[27] Smith offers other arguments for thinking of reasons as Kantian. One of them is that we cannot derive normative reasons from desires that are arbitrary. This type of argument is discussed in Chapter 2.

> Suppose someone tells me she has a reason to take a holiday and that I think I would have no reason to take a holiday in the circumstances she faces. Provided we have taken proper account of the *de se* considerations that might be relevant to her choice, and provided that we have taken proper account of the way in which her preferences may constitute a relevant feature of her circumstances, it seems that I can straightforwardly disagree with her about the rational justifiability of her taking a holiday in the circumstances she faces, a disagreement I can express by saying 'She thinks that there is a reason to take a holiday in her circumstances, but there is no such reason'. (*Ibid*., pp. 171-172.)

I'll call the person who thinks she has a good reason to take a holiday Jane, her opponent – Bob, and use the holiday example to go through Smith's argument (*Ibid*., pp. 166-167). According to Humeans, when Jane says she has a good (i.e. normative) reason to go on holiday, she is talking about normative reasons-relative-to-*Jane's*-desires. When Bob says that Jane has no good reason to go on holiday, he is talking about normative reasons-relative-to-*Bob's*-desires. Thus, if we accept that normative reasons are relative to our desires, Bob and Jane are talking about different things, and can't disagree. But we ordinarily take them to disagree – we take them to be talking about the same thing: a non-relative reason. So, our ordinary concept of a normative reason is a non-relative (anti-Humean) one.

We should also note the two provisos in the quotation that describes the holiday example above: before disagreeing about normative reasons, we need to take care of *de se* considerations and the agent's personal preferences. This is because such considerations and preferences may look relativizing, but actually aren't. Let us go through an example of each.

Seemingly relativizing case 1: de se considerations

A reason 'to save my child from drowning' is not available to me when someone else's child is drowning. Even so, this reason is not relative. In the right circumstances (if my child were drowning), this reason would be available to me, and it would justify *simpliciter* my action.

Seemingly relativizing case 2: personal preferences

If I say 'I like wine, and that's a good reason for me to go to the bar', and you answer

that it's not a good reason for you, you are not relativizing my reason. Rather, the reason is tied to my particular circumstances, which include my psychology, and you agree that if you were in the same circumstances, including my psychology (i.e. if you liked wine), then you also would have a reason to go to the bar. Once all circumstances have been taken into account, my reason is not relative to me. (*Ibid*., pp. 168-171.) (This, one cannot help noting, sounds very much like the NRF, and, as we have seen above, accepting this constraint on normative reasons does not require us to go beyond the level of the sensible knave.)

In response to Smith's argument, Sobel (1999) shows that Bob can be talking about reasons-relative-to-Jane's-desires as well. According to Sobel, if Bob says that Jane has no reason to go on holiday, he means that Jane would not be motivated to go on holiday had she deliberated correctly.[28] There is a fact of the matter what Jane would be motivated to do. If Jane would be motivated to stay after correct deliberation, then Bob is right, and she does not have a good reason to go (*Ibid.*, pp. 143-144).[29]

Sobel's response is perfectly adequate to Smith's argument as stated. Indeed, this is what Humeans should say about disagreement. But maybe Smith's point is better interpreted in a different way. Ordinarily, we think that there can be two types of disagreement about normative reasons:

> a) disagreement about whether one should be taking this particular means to the end they pursue
>
> and
>
> b) disagreement about whether one should pursue a particular end.

---

[28] This response uses Williams' (1979) thesis that nothing counts as a reason unless an agent can be motivated by it (internalism about normative reasons). According to this thesis, if Jane is not motivated to stay at home rather than take a holiday after correct deliberation, she has no reason to stay at home. There are no reasons that a fully rational agent is insensitive to.

[29] As pointed out by Prof. Pink, one can easily see the truth of this point when one is required to give advice. If Jane asks Bob for advice about what she should do, then of course he should take Jane's motivations into account: if he fails to do so he is either incompetent or immoral. But advice situations are different from situations of disagreement. I'll use Dancy's example to illustrate (2000, pp. 34-35). Suppose Jane has a project of soundproofing her house so that not even the slightest noise penetrates from the outside; Bob thinks it is a silly idea, but he is knowledgeable about different types of insulation, and knows that, given what Jane is trying to do, fibreglass insulation is the best choice. Armed with this information, Bob can give Jane good advice, even though he still thinks there is no good reason for her to soundproof her house. So, giving advice about the best means of achieving a goal clearly involves taking into account the motivations of those who seek advice.

Humeans, however, can accommodate only the first type of disagreement, and Sobel's response highlights this. Suppose Jane thinks that she should take a holiday because her work has not been going well lately, and she thinks that a holiday will relieve her stress and enable her to solve the problem she is facing on her return. (So, Jane's holiday is a means to doing her job well.) Bob, however, thinks that Jane is on the brink of solving the problem that is plaguing her, and if she goes on holiday now, she will lose her momentum. Jane and Bob disagree about the best means to the end Jane pursues. However, they cannot genuinely disagree about whether Jane should pursue a particular end, say, doing her job well. If Bob thinks that Jane should not pursue the type of work that she does, yet he knows that Jane wants no other work, and this desire is not based on misinformation, etc., then Bob, on a traditional Humean account, cannot be justified in claiming that Jane should not pursue her work. But, as our pre-theoretical commitments show, we do not think that someone has a good reason to do something just because they want to do it. Nor do we think that someone has no good reason to do something just because they don't want to do it. Thus, we ordinarily suppose that there is more distance between desires and normative reasons than traditional Humeans allow.

One way to create such a distance is to admit that sentiments do not provide normative reasons directly, but by enabling mastery of evaluative concepts. Such concepts then figure in one's evaluations, which, depending on what metaphysics of value one has, either constitute one's good reasons (if values are not real) or represent good reasons (if values are real). I defend this account in the following chapters, so for now I shall only show how it applies to the debate between Sobel and Smith. If values are real, then we can easily accommodate genuine disagreement about ends: either Bob's evaluation of Jane's end ('Jane's work is worthless.') or Jane's evaluation of her end ('My work is worthy of pursuit.') is correct. What happens is values are not real? Well, if both disputants agreed that values are not real, they would be unlikely to criticize each other's ends, so the dispute would not even arise. This would include abandoning some common-sense commitments, since we ordinarily accept value realism when we argue about what one has a good reason to do: I would not ordinarily say 'You have no good reason to go on holiday.' unless I thought that there really were such things as good reasons. However, even if values are not real, we can still accommodate the point that

Smith' argument is bringing out, i.e. the point that traditional Humean theories tie normative reasons to desires too closely. This is because, on my account, desires, or, more broadly, sentiments, no longer have a direct connection to normative reasons. Sentiments only enable mastery of evaluative concepts, which figure in evaluations, and evaluations, on an irrealist metaphysics, constitute one's normative reasons. Thus, evaluation and motivation can come apart. Jane can evaluate something as good, and hence have a normative reason to pursue it, without being motivated to pursue it. So, even if we accept an irrealist metaphysics of values, Bob can be justified in saying that Jane has a normative reason to do something (say, stay at home), even though she is not motivated to do it, as long as Jane positively evaluates staying at home.

One may ask how we can put this distance between motivations and evaluations and remain Humean. I spell this out in the next chapter, but here is a preview. The position proposed here is (weakly) Humean because I accept that sentiments are still necessary for having (access to) normative reasons. What I reject is a traditional explanation of why they are necessary. On a traditional Humean account, this necessity is explained in terms of promotion: I have a normative reason to do whatever promotes my desires. On my account, it is explained in terms of mastery of evaluative concepts.

3. Conclusion

In the first part of this chapter, I have shown that a formal version of a rationalist theory fails to provide evaluations, which are necessary for acting for good reasons. In the second part, I have explained the appeal of Kantianism: it tries to accommodate our intuition that there is nothing special about me as user of reasons, as expressed by the Normative Reason Feature. I have also showed that the possibility of disagreement about normative reasons points to a problem for traditional Humean theories. The problem is that such theories do not distinguish between motivations and normative reasons, as will be made clearer in the following chapter.

Chapter 2. Sentimentalism

1. Introduction

The aim of this chapter is to explain what I consider to be the main problem for a sentimentalist theory and provide my solution to it. My example of a sentimentalist theory, as in the previous chapter, will be Humeanism. Traditional Humeanism fails to explain how desires, which are *prime facie* non-normative, generate normative reasons to do what I want. This problem appears in the literature in different guises. Sometimes it is pointed out that Humeans cannot explain why some desires intuitively fail to provide normative reasons. Sometimes criticisms fall on the Humean conception of instrumental rationality and their inability to explain what provides me with a reason to satisfy my desires. In this chapter, I shall concentrate on the first of these, as I think it is the most basic formulation of the problem.

I argue that the problem arises because traditional Humeans explain why desires are sources of normative reasons in terms of promotion. They say that I have a normative reason to do whatever promotes my desires. I offer an alternative theory, according to which desires are sources of normative reasons for action because they are necessary for mastery of evaluative concepts.

2. The Humean position

Humeans claim that all good reasons depend on our desires. I think the best way to understand this claim is to say that Humeans are sceptical about the extent to which reasoning can change our desires. In what follows I explain how traditional Humeans see the role of reasoning by using Williams' (1981) and Blackburn's (1998, pp. 238-269) accounts.

Humeans accept that reasoning alone can *help* us to act well. First, it provides information that is relevant to achieving our aims. For example, one wants a gin and tonic, and believes this is what's in the glass. But one is wrong: the glass contains petrol. In this case, one has no normative reason to drink what's in the glass, and it's obvious why – because one's belief that the glass contains gin and tonic is false.

Reasoning can also refine our desires in other ways – e.g. by one's thinking through what it would be like to do what one wants, or what the consequences of doing it would be. Hence, if we want to act on good reasons, we should try to minimize malfunctions of reasoning, such as memory lapses, inattention, etc. Secondly, reasoning helps desire-satisfaction. It does so by requiring us to take means to our ends and alerting us to the existence of incompatible desires. For example, if I want to listen to music, I should put the player on. If I want to maximally satisfy my desires, reasoning requires balancing them. For example, I want to get new shoes and I want to go on holiday. It may be that these desires are in conflict: if I bought the shoes, I would not have enough money left for the holiday. Reasoning alerts us to such inconsistencies, and may help us find the way to maximally satisfy our desires (for example, if I set the money aside now and go on holiday to Poland, shoes will be cheaper there, so I might end up both going on holiday and buying the shoes). If no such solution is forthcoming, reasoning alerts me to the fact that I have to choose satisfaction of one desire over another.

But Humeans circumscribe the role of reasoning. Reasoning alone can't select ends and it can't provide criticism of ends. If there are no rationally required ends, it follows that whatever ends we have, they cannot be rationally criticized: if I fail to take means to some end, I am irrational (taking means to ends is rationally required), but if I fail to have a particular end, I am not (since all ends are rationally optional). It is important to note what does not follow from the fact that reasoning alone never selects ends. It does not follow that ends cannot be criticized *at all*. Only one type of criticism – rational criticism – is ruled out. We can still criticize ends – our own and others' – in the light of our desires and concerns. In other words, we evaluate ends, and that, as we have seen in the first chapter, is not something that reasoning provides. For example, I may think that the pleasure of buying shoes is quite trivial in comparison to knowledge and open-mindedness that travelling can bring: I can criticize wanting to buy shoes as being too materialistic, but not irrational.

There are passages where Humeans and their opponents sound surprisingly similar when describing the role of desires in deliberation. Here is an example:

> In practical reasoning, we need not only to select among possible plans of action but also to select among considerations to be taken into account in deciding what to do. … So it may be that the person who “has a desire for coffee ice cream” …

> has taken the desirability of having ice cream as one of the considerations to be taken into account in deciding what to do in the near future. It is, so to speak, "on her deliberative agenda", whether or not she ultimately forms an intention to act on it or not. (Scanlon 1998, pp. 46-47.)

> Typically, in deliberation what I pay attention to are the relevant *features* of the external world: the cost of alternatives, the quality of food, the durability of the cloth, the fact that I made a promise. I don't *also* pay attention to my own desires … My own concerns and dispositions determine which feature I notice and how I react to them. If I am a miser, the cost takes my attention; if I am a gourmet, the quality of the food does; if I am prudent, the durability of the cloth; if I am not a knave, the fact of the promise. (Blackburn 1998, pp. 253-254, italics in original.)

Both Blackburn and Scanlon agree that we deliberate about features of the world. Both agree that desires single out the features we consider salient in deliberation. The difference is in the order of explanation. For Blackburn, the existence of reasons is dependent on my having a desire: I only treat R_1, R_2 and R_3 as reasons because I have a desire. For Scanlon, reasons exist first, and desire comes about as a result of endorsing them. Consequently, for Blackburn, deliberation is only possible in the light of some desire or another. I can stand back from any particular desire – say, I may wonder whether I should really be a gourmet, and whether I should pay so much for a meal when I could buy something nice for my family instead. But this evaluation is only possible because I have other concerns, such as concern for my family.

For Blackburn, there necessarily is a blind spot in deliberation. This blind spot is the standpoint from which I currently evaluate other concerns. The concern from which I evaluate other concerns may change, so, in the end, I can survey the whole field, but not at the same time. Scanlon would consider the existence of such a blind spot as skewing deliberation. For each decision, such as the decision to go to the restaurant or to buy ice-cream, there are reasons. And these reasons can be evaluated independently of which desires I seek to satisfy; in Backburn's terms, they can be evaluated independently of any of my concerns. Scanlon argues that desires are not the starting points of practical deliberation (Scanlon 1998, p. 43). Blackburn would agree to some extent. It is true of any desire taken individually that it does not have to be such a starting point. But it does not follow that the starting point of a deliberation is not some desire or another. Desires, for Blackburn, are the only standpoints possible.

3. Problems for Humeans

As we have seen above, Humeans provide refinements as to which desires would count as normative reasons: I have to have relevant true beliefs, no relevant false beliefs, I have to balance my ends and take the means to them, etc. But rationalists argue that this is not enough to account for the force of normative reasons, as the following problems show.

Problem 1: Counter-examples: too many and too few reasons

Too many: having a desire to φ *does not automatically provide a good reason to* φ (e.g. Quinn 1993, Dancy 2000). At least some of my desires don't provide me with a normative reason to do what I want. If I want something silly, then my wanting to do it does not make it any less silly; in other words, my wanting does not give me a good reason to do it. Cutting my finger off is (in normal circumstances) a bad idea. And if I develop a liking for it, this does not change the situation in the slightest. This is what Schroeder (2007) calls the 'Too Many Reasons Problem'.[30, 31]

Too few: lacking a desire to φ *does not mean one has no good reason to* φ. There may be some things that anyone has a reason to do even though they don't want to. Moral reasons, for example, seem to be like that: 'if murder is genuinely wrong, then there must be some reason for anyone not to murder people' (Schroeder 2007, p. 103). Since Humeans claim that what we have reason to do depends on what promotes our desires, Humeans are unable to explain such reasons, unless they can show that there are some desires that any agent necessarily has. This is what Schroeder (2007) calls the 'Too Few Reasons Problem'.

Problem 2: Instrumental reasoning (e.g. Quinn 1993, Korsgaard 1997, Dancy 2000). Humeans hold that we are rationally required to take means to our ends: if I want to

[30] I picked an unusual desire and gave no good reasons for it, because this is how traditional Humeans must think of at least some desires. Admitting that *all* desires are held for normative reasons is admitting that reasons generate desires, which is tantamount to abandoning traditional Humeanism. So, traditional Humeans must construe some desires as ones not held for reasons, as inexplicable urges. And these urges must, according to traditional Humeans, generate normative reasons. (Quinn 1993)

[31] I should note, following Hampshire (1999), that sometimes it is perfectly acceptable to cite your desire as a good reason for action. For example, the fact that I really like this particular painting often gives me a good reason to buy it. Yet, there are problematic cases, in which my desire gives me no good reason to do as want.

listen to music, I should put the player on. They also hold that none of the ends we have are rationally required: it's not the case that rationality requires me to want to listen to music. But how does a non-obligatory end create an obligation to take the means? Instrumental reasoning can transfer normative force from ends to means, but it cannot create normativity from nowhere.[32]

Problem 3: A reason to satisfy my desires (e.g. Nagel 1970, Quinn 1993). Humeans assume that I have a reason to satisfy my desires – if I want to listen to music, I have a reason to put the player on. Why do I have this reason? Well, because putting a player on will satisfy my desire for music. If there were no relationship between putting it on and the satisfaction of my desire, then I would have no reason to do it. So, in order for a Humean theory to work, we need to assume that I have a reason to satisfy my desires. Humeans hold that any reason is explained by a desire. But which desire explains a reason to satisfy my desires? If this reason is not explained by a desire, then the Humean theory is false. So, a Humean must say that the reason to satisfy my desires is explained by some desire I have. A desire to satisfy my desires seems an obvious candidate. But suppose that I lack this desire. Then, if I have a desire to listen to music, I have no reason to put the player on. This seems wrong – surely, for Humeans, the desire to listen to music must be capable of generating reasons on its own, without the help of a further desire to satisfy my desires. Besides, if in lacking this desire I lack reasons altogether (i.e. if I stop being an agent), then it looks like this desire is a rationally required one. So Humeanism is false, because it denies that any rationally required desires exist. (N.B. A Humean may say that desire is a psychological state that disposes one to seek its satisfaction. This may be true, but it fails to give desires any normative force, i.e. fails to explain how desires can provide normative reasons.)[33]

---

[32] One could deny that this is a problem for Humeans. As Lillehammer (2007, Ch.3) argues, Humeans can admit instrumental rationality and retain what is distinctive about their view. The battle between Humeans and their opponents is not about whether there are any norms of rationality *at all*, but about whether there are any substantive, action-guiding ones. I am sympathetic to this understanding of the debate, and it is part of the reason why I concentrate on the more basic problem of counter-examples.

[33] As pointed out to me by Prof. Pink, there are two possible Humean views, that are often run together both by Humeans and their opponents. According to the first view, there is no such thing as the good, and one's desires are evaluated in terms of consistency with each other, but not in terms of what they aim at. This is a sort of coherence theory of the good. According to the second view, good is whatever the agent wants. This second form of Humeanism does not face the problem about instrumental reasoning and the problem of finding a reason to satisfy my desires: if my wanting something makes it good, then this goodness of the end transmits itself to the means; it also gives me a reason to satisfy my desires, because whatever I want is good.
Both versions of the view, however, face the problem of counter-examples, and that is the problem I concentrate on in the main text. According to the first, coherentist, view, my desire gives me a reason to do something as long as it coheres with other desires I have. But, as we shall see in what follows, even if what I want coheres with other things, it does not always, intuitively, give me a good reason to

In what follows, I shall concentrate on the Too Many Reasons problem, but before examining Humean responses to it, I shall sketch a rationalist alternative, so that we can see clearly what the problem for Humeanism is and what we have to aim for in our solution.

3.1. A rationalist alternative

Anti-Humeans solve all the problems presented above by claiming that desires must be explained in terms of reasons, not vice versa. Several authors gave such an explanation: Quinn (1993), Scanlon (1998), Dancy (2000). I shall concentrate on Scanlon's version (1998, pp. 37-49). The main force of Scanlon's argument comes from the thought that 'having what is generally called a desire involves a tendency to see something as a reason' (1998, p. 39.). He argues for this claim in the following manner.

Scanlon's argument

I. Desires don't provide sources of motivation independently of reasons.
II. Desires don't provide sources of justification independently of reasons.
III. Therefore, desires are to be analysed in terms of reasons.
IV. Humeans claim that reasons are to be analysed in terms of desires.
V. III and IV make Humeanism circular.[34, 35]

The first premise is supported by the following considerations. Suppose I am motivated to act. There are two things we can mean by that, says Scanlon. I may be motivated because I have an urge to act. 'An urge' here means that I don't see anything good about the action. Scanlon dismisses this, saying that such urges are not what we ordinarily

do it, so the coherentist view faces the Too Many Reasons problem. There are also cases in which doing *x* does not cohere with my other desires, yet, intuitively, I have a good reason to do *x*. E.g. I want to spend all my money on myself rather than support my young child. This desire coheres better with other things I want, yet, intuitively, I still have a reason to support my child. So, the coherentist view faces the Too Few Reasons problem. According to the second view, good is whatever I want. But just because I want something, it does not automatically make it good (Too Many Reasons problem), and just because I don't want something, it does not mean it is not good (Too Few Reasons problem).

[34] IV and V are not stated in Scanlon's text.

[35] Korsagaard (1997) provides a similar argument against Humeanism. She argues that if desires are non-normative, then they cannot be sources of normative reasons. But if desires are normative, and this normativity derives from the principles of reason, then Humeanism is circular.

mean by 'desire'. This may be true, but a Humean is not tied to using an ordinary notion of desire, so we need to say more in order to exclude this option of desire providing a motivation to act. We can exclude it for reasons that Dancy (2000, p. 36) proposes: doing something when I see no reason whatsoever to do it is surely a pathological, not paradigm, example of motivation. The second thing we may mean by saying that I am motivated to act is that I have a desire in the directed-attention sense:

> [a] person has a desire in the directed-attention sense that P if the thought that P keeps occurring to him or her in a favourable light, that is to say, if the person's attention is directed insistently toward considerations that present themselves as counting in favour of P. (Scanlon 1998, p. 39.)

It is clear that I can be motivated without having a desire in the directed-attention sense – for example, when I am about to drink a foul-tasting medicine. Thus, directed-attention desires are not necessary for motivation. In addition, when I am motived by a desire in the directed-attention sense, I am motivated by reasons that my attention is drawn to, not by the desire itself. For example, if I want an ice-cream, my attention is drawn to its pleasurable cool taste. It is this expected pleasure that motivates me; there is no element of wanting distinct from this perception of pleasure as a reason. So, desires do not provide sources of motivation that are independent of seeing things as reasons.

Scanlon argues for the second premise in the following manner. Suppose I have a good reason to do something because, as we say, I want to do it. There are several things we may mean by this. First, I may have a good reason to buy an ice-cream because I think that its cool, pleasant taste provides me with a good reason to buy it. Here the expectation of pleasure provides a normative reason. But desire does not provide an independent reason; rather, it is 'an endorsement of a reason', as Raz puts it (1986, p. 141). To see this clearly, suppose you wanted an ice-cream whilst not expecting any pleasure from its taste;[36] then it's difficult to see why you have a good reason to buy it. Secondly, I may have a desire in the directed-attention sense. My mind keeps returning to all the things that make the object of desire attractive. For example, if I want to buy a new computer, I keep thinking about how nice the big new screen would be. As in the case of expected enjoyment above, the justificatory work here is done by the features I take to be reasons. My desire may be an expression of my positive evaluation of these

[36] Suppose also that you don't have any other reason to eat it apart from the pleasant taste.

features, but without them it lacks justificatory force. Moreover, making these positive evaluations does not automatically give me a good reason to buy a new computer. I may have my attention insistently directed at how nice it would be to buy a new computer without for a moment thinking that I have a good reason to do so. Thirdly, I may have an intention. Intentions can give you reasons that you would not have had otherwise. For example, if I decided to buy a computer for reasons R_1, R_2 and R_3, and I have no reason to reconsider, then I have a reason to go to the computer shop. If I am as yet undecided, I have no reason to go to the computer shop. But intentions are not independent sources of reasons. My intention is just a place-holder for the three original reasons I had: my intention justifies my action just to the extent that R_1, R_2 and R_3 justify it. Desires, for traditional Humeans, are meant to be different: they provide reasons for action where originally there were none. Thus, in none of the cases considered do desires provide a source of justification that is independent of reasons.

The notion of desire identified by Scanlon – desire in the directed-attention sense – captures our ordinary notion well.[37]

- I can do what I have no desire, in this sense, to do, as when I drink a foul-tasting medicine. I take the medicine without my attention being grabbed by its attractive features.
- Desires, even when explained in terms of reasons, may resist one's considered judgement. For example, even if I know I have no reason to buy a computer, my attention may still turn to the attractiveness of getting one.
- It is true to the phenomenology of desires: when we want something our attention is captured by the features of the desired object.

If desires are explained in terms of reasons, then we face none of the problems presented above:

Problem 1.

Too many reasons

Having a silly desire, i.e. a desire that I have no (all things considered) reason to have, is compatible with having no good reason to do what I want. This is because, as noted

[37] N.B. A Humean may use 'desire' not as an ordinary notion, but as a technical term. Hence the considerations listed below are further support for Scanlon's understanding of desires, not part of his main argument against Humeanism.

above, desires can persist in the face of contrary reason-judgements.

Too few reasons

We no longer face the Too Few Reasons problem, because normative reasons do not depend on the agent's desires: she may have a reason to do something even though she does not want to.

Problem 2. Instrumental reasoning. I have a reason to take means to my end only if I also have a reason to pursue this end. For example, it is only if I have a reason to listen to music (e.g. because I expect to enjoy it) that I have a reason to take a means to doing so. Instrumental reasoning does not create its own normativity, but the reasons I have for the end are transferred to the means.

Problem 3. A reason to satisfy my desires. We no longer need to look for a reason to satisfy my desires because they, contrary to the traditional Humean assumption, do not always create reasons to satisfy them. If desires are explained in terms of reasons, I want *x* for reasons R_1, R_2 and R_3, and I may have (or fail to have) a reason to satisfy my desire for *x* depending on how good (or how poor) R_1, R_2 and R_3 are.

So, rationalists say, desires are not independent sources of motivation or justification. They depend on reasons for that. Therefore, desires are to be analysed in terms of reasons – a claim that Humeans must reject on pain of circularity. How do Humeans respond?

4. Humean responses

I shall concentrate on the problem of counter-examples, especially on the Too Many Reasons Problem, as it seems to be the basic one. The other two[38] problems – about instrumental rationality and about a reason to satisfy my desires – can be derived from the fact that desires don't automatically give me good reasons for action.

The Too Many Reasons problem is not easily tackled – it looks like any Humean theory will have to insist that in some special cases, we have a good reason to do what we want, even if we want something silly. A Humean theory creates good reasons in cases

[38] I do not include the Too Few Reasons problem here as it is, together with the Too Many Reasons problem, subsumed under the problem of counter-examples.

where intuitively there aren't any. Yet,

> to the extent that a Humean is willing to admit to accepting results that are intuitively false, other philosophers are going to legitimately infer that he has simply changed the subject, and is talking about something else entirely. (Schroeder 2007, p. 86.)

The Too Many Reasons problem is generated by two claims:

Intuition: silly desires do not provide one with normative reasons.
Traditional Humean claim: desires provide one with normative reasons to do whatever promotes them.[39]
There does not seem to be an obvious way of excluding silly desires from the Traditional Humean claim.[40] Still, there are several things one can do.

- Reject the *Intuition*. This option would fall prey to the charge of changing the topic. If a Humean says that our intuitions about normative reasons are unreliable, then it does look like what Humeans are talking about is not what we ordinarily talk about when we say we have a good reason to do something.
- Explain why the *Traditional Humean claim* does not apply to silly desires.
- Reject the *Traditional Humean claim*. If this option is taken, it may be difficult to say why one's theory is still Humean. This is the option I favour. The resulting theory is certainly a departure from traditional Humeanism, but is

[39] How the promotion relation is to be construed is a matter of debate (Schroeder 2007, pp. 110-113). One way to understand it as an identity relation – I have a reason to do A only if I want A. This, as Schroeder points out, is too strong: Ronnie may have a reason to go to the party because he wants to dance, and there will be dancing at the party. The action – going to the party – is not the same as the desire it promotes – wanting to dance. Schroeder weakens the promotion relation: I have a reason to do A only if doing A increases the likelihood of the object of my desire obtaining. This, however, leads to intuitively unacceptable consequences, as Schroeder himself admits. For example, eating my car increases the likelihood of my getting my daily dose of iron. I want to get my required dose of iron, so I have a reason to eat my car.

[40] Williams' (1979) considerations do not help here. According to Williams, my desires will only give me good reasons if they (desires) satisfy the following criteria:
1. My desire is not based on a false belief.
2. I have relevant true beliefs (e.g. I have true beliefs that will help me satisfy my desire).
3. My desire is refined by the process of deliberation in terms of being consistent with my other desires and means to their satisfaction.
A silly desire can satisfy all these criteria. If you have doubts about the finger-cutting example, I'll use another one, that comes from Schroeder (2007). Aunt Margaret wants to reconstruct a scene from a furniture catalogue on Mars, for which she needs a spacecraft. No one will give her a spacecraft, so she has to build one herself. According to Humeanism, she has a good reason to build a spacecraft. Her desire does not obviously fail any of Williams' constraints, but it is still a silly desire that should not, intuitively, give rise to a normative reason.

inspired by it: it holds onto the claim that sentiments are necessary for having (access to) normative reasons.

4.1. Traditional Humean response

There are three different ways of explaining why the *Traditional Humean claim* does not apply to silly desires. The first one is to point out that silly desires are outweighed by other desires I have. This is the option I shall concentrate on in detail, since I think it is the most viable strategy for traditional Humeans, but I mention the other two briefly. The second way of solving the Too Many Reasons problem is to say that there are two different types of desires, and that only desires of one type, but not another, provide one with normative reasons. Frankfurt (1971), for example, distinguishes between first- and second-order desires. He argues that my first-order desire fails to give me a normative reason to promote it if I don't have a corresponding second-order desire. My desire to cut my finger off, for example, does not give me a good reason to do it if I don't want to have this desire. My second-order desires correspond to my values, or things that I have reasons to do. The are two general problems with this sort of response.[41] First, as Quinn (1993) points out, traditional Humeans have to think of at least some desires as inexplicable urges (or, in Frankfurt's terminology, as desires lacking corresponding second-order desires) which still provide me with normative reasons. This is because traditional Humeans hold that desires explain reasons, but not vice versa. If all my desires that give me normative reasons do so because they are held for reasons, Humeanism is circular. At this point one may say that Frankfurt does not mean that second-order desires are desires that are held for good reasons; rather, it is just a brute fact that they provide us with normative reasons whilst first-order desires do not. But then it is difficult to see why second- rather than first-, or *n*-order, desires should be identified with what I value. It opens up a possibility for having sensible first-order desires and silly second-order ones.[42] The second problem for the theories that divide

[41] I discuss this response by using Frankfurt (1971) as an an example, but the same problems face similar theories developed by other authors. (Schroeder (2007, p. 85, n. 3) mentions Williams (1973) and Watson (1975) as proponents of the two types of desire strategy. I am not entirely convinced that they are. Williams is not trying to solve the Too Many Reasons problem; he targets utilitarians' inability to make space for personal projects. This does not commit Williams to saying that only desires that promote my life projects give me good reasons. As for Watson, it is not clear that he is a Humean at all, since he admits that value judgements may provide a source of motivation that is independent of desires.)

[42] This is indeed the criticism made of Frankfurt's view by Watson (1975).

desires into two qualitatively different types is noted by Schroeder (2007, pp. 85-86). Such theories, on the one hand, still allow some silly desires to generate normative reasons. For example, it is possible that I may want to cut my finger off and desire to have this desire. On the other hand, such theories exclude some sensible desires from generating normative reasons. For example, if I want to go to the shops, which, let us suppose, is a sensible thing to do in the circumstances, it is possible that I do not have a second-order desire to do so. Nor do I normally form second-order desires for every sensible first-order desire I have.[43]

One could also try to adapt the distinction between wide- and narrow-scope rational requirements as a solution to the Too Many Reasons problem.[44] Suppose that getting a knife from the kitchen is a necessary means for cutting my finger off. So, if I want to cut my finger off, I ought to get the knife from the kitchen. This last proposition is ambiguous. It can be read as ought having a wide scope, i.e. modifying the whole proposition: I ought (if I want to cut my finger off, to get the knife from the kitchen). Alternatively, it can be read as ought modifying only the consequent: If I want to cut my finger off, I ought (to get the knife from the kitchen). Blackburn (1998, pp. 242-243) and Broome (1999) argue for the first, wide-scope reading. On such a reading, I am rationally required to either take the means to my crazy end or to abandon the end. This may be a way of dealing with silly desires. If oughts are wide-scope in this context, then abandoning the desire, rather than promoting it, is an option I have. And this is what I should do in the case of a silly desire, although not in the case of a sensible one. The problem with this response is obvious: which desires should I abandon, and which ones should I promote? In order to answer this question, we need a substantive distinction between silly and sensible desires. In which case this response collapses into the previous one, which tries to distinguish between different types of desires, and the problems for that have already been discussed.

Having briefly considered these two responses, I shall now concentrate on the one that

[43] One may object that I may have an implicit second-order order desire in such cases, and it can be made explicit by questioning. However, I still think it is psychologically unrealistic to claim that I form an (implicit) second-order desire for every sensible desire I have. Nor can Frankfurt explain why this should be the case, given that there is no qualitative difference between desires of different orders.

[44] The defenders of such a theory disagree about what these rational requirements are. Blackburn (1998, pp. 242-243) talks about 'oughts', Broome (1999) says it is a requirement that is distinct from both 'reasons' and 'oughts' relations. In what follows, I shall talk about oughts, but the debate about what rational requirements are does not affect my argument.

seems most developed and most promising. This is the idea that silly desires fail to provide normative reasons to satisfy them because they stop me from satisfying my other desires. If I cut my finger off, for example, I would not be able to satisfy any of my desires that require a full compliment of fingers. This particular desire does generate a reason, but it is outweighed by other desire-generated reasons I have. Any of my desires can be evaluated from the point of view of other things I want. E.g. if I want to play the piano, then this desire may well outweigh my desire to cut my finger off. Such outweighed desires fail to give me normative reasons, even though they would, according to Humeans, give rise to normative reasons where they not outweighed.

So, according to this response the *Traditional Humean claim* that generates the Too Many Reasons problem does not apply to silly desires because they are outweighed by other desires one has. There are two problems with this solution. The first problem is that this response has failed to get rid of the original counter-example.

Silly desires still generate good reasons

This response says that although the silly desire to cut my finger off is outweighed by other things I want, there nothing wrong with its reason-producing powers *per se*. Were other desires absent, this desire would give me a good reason to satisfy it. In this case, Humeans still say that I do have a normative reason to do something stupid. Someone who already has a strong Humean intuition (and I do) will be happy to agree that, since my desire is outweighed by other desires, it does not give me a good reason to do as I want. But anti-Humeans would not be satisfied with this line of thought; for them, it just pinpoints the problem rather than provides a solution. If cutting my finger off was a silly idea to start with, why should it make any difference to its silliness whether I have desires that outweigh it? The result anti-Humeans want is that my silly desire fails to provide me with a reason *because* it is silly, not because there are some other desires that would not be satisfied because of it.

A Humean may respond that we only call a desire 'silly' when it puts one's other desires into jeopardy. There is nothing wrong with a desire *per se*, it is only silly against the background of other desires.[45] If competing desires are removed, my previously silly desire would give me a normative reason to satisfy it. Thus, the Humean theory gives us the correct result. This response does not work. To see why, consider a variation on a

[45] Blackburn says something along those lines: my desires can only be evaluated from the point of view of my other desires. So, if I only have a single desire, it cannot be evaluated (1998, p. 240).

famous thought-experiment (Williams 1973). I am offered to kill one of ten prisoners. If I do this, the others will be unharmed. If I refuse, then all ten will be killed. In this variant, I have the option of sacrificing myself instead of killing one of the prisoners: if I die, all ten prisoners will be unharmed. Suppose I have a desire to sacrifice myself in these circumstances.[46] It is very clear that this desire puts most, if not all, of my other desires in jeopardy. There are a lot of competing considerations, I am not someone who will give my life gladly and without regrets. Suppose my desire to sacrifice myself is outweighed by my desire to live. So, according to the traditional Humean response, my desire to sacrifice myself is a silly one and fails to generate a normative reason. But even someone who says that I should not sacrifice myself in these circumstances, for example, because the demand is too great, would not say that this desire is a silly one, of the sort that fails to generate a normative reason. Contrary to the Humean contention, some desires that are outweighed are not silly. Conversely, some desires that are not outweighed can still be silly. We would ordinarily say that my desire to cut my finger off fails to provide me with a good reason to satisfy it even before we are told of competing considerations, if any. This latter point can be better brought out by making the example more detailed. Suppose that I could help a nearby child, which requires a full complement of fingers. I am aware that the child needs help and that I must have my fingers intact to help her, yet I do not want to help. In this case, my desire to cut my finger off is not outweighed, but it still fails to provide me with a good reason to do as I want.

Reasons' weight

The traditional Humean response says that some desires provide no good reasons to satisfy them because they are outweighed by other desires. In this case, a Humean owes us an account of the reasons' weight. Traditional Humeans account for weight of reasons in terms of strength of desire: R_1 is weightier than R_2 iff I want R_1 more. But this account of reasons' weight presents at least two problems.

- If I want to cut my finger off more than I want to do anything else, then, according to this method of weighting up reasons, I have a good reason to cut my finger off. This generates a version of the original counter-example. A Humean says that just because I want something silly *very much* I have a normative reason to do it. But ordinarily we think that strength of desire makes

[46] Some may say that we should not talk about my desires here, but I assume that sacrificing myself is driven by a desire to do so in order not to beg the question against Humeans.

no difference here. So, either a Humean is incorrect, and silly but strong desires provide no normative reasons, or she is changing the subject and is no longer talking about good reasons as we ordinarily understand them.

- Anti-Humeans may justifiably complain that this is not an account of a *normative* reason's weight. The strength of desires may be a good measure for motivating reasons, but how good a reason I have to do something just does not depend on how much I want to do it.

So, the traditional Humean solution fails to explain why silly desires don't generate normative reasons.

4.2. Schroeder's response

The Too Many Reasons problem has often been taken lightly by Humeans. Schroeder (2007) is an exception: he both recognizes that it is a problem, and that it requires abandoning some traditional Humean assumptions. Schroeder's solution is partially to reject the *Intuition* and to provide an account of reasons' weight that excludes silly desires from the *Traditional Humean claim*. Rejecting the *Intuition* outright will fall prey to the charge of changing the topic, but Schroeder tries to remedy this by making all good reasons weighty. On the one hand, he accepts that all desires – even silly ones – generate normative reasons. He thinks that, for example, I have a reason to eat my car, because doing so will get me my daily portion of iron. He admits that it is not a weighty reason, but nonetheless it qualifies as a reason (pp. 95-96). On the other hand, Schroeder rescues common-sense talk about normative reasons by holding that what we ordinarily mean by a 'good reason' is a 'weighty reason'. And then he proposes an account of normative reasons' weight that purports to show that all reasons generated by silly desires are not weighty.

So, Schroeder's solution to the Too Many Reasons problem depends on his account of reasons' weight (2007, pp. 123-145). He rejects the account of normative reasons' weight in terms of desire strength. Instead, Schroeder says, one reason is weightier than another if I have more reason to place weight on it. This is not (quite) as trivial as it seems. For example,[47] suppose Ronnie is deciding whether to go to the party. He has a

[47] Adapted from Schroeder 2007, pp. 127-128, pp. 132-133.

reason to go – R_1 (there'll be good food). He has a reason not to go – R_2 (his ex-girlfriend Isabel will be there). Now suppose he knows that Isabel has a new boyfriend who lives in another city. If so, it is just as likely that Isabel will visit him instead of turning up at the party, and this consideration (R_3) is a reason to place less weight onto R_2. On the other hand, Isabel's new boyfriend may want to meet Isabel's friends; this undercuts consideration R_3, etc. At some point, Schroeder insists, these defeaters run out, so all we have to do is 'go back through the chain' (*Ibid*., p. 137) and determine the weight of the original reasons – in this case, R_1 and R_2, by seeing which one of them is not undercut.

This analysis, unlike the traditional Humean account of normative reasons' weight, makes weight depend on something that is normative – on other reasons. Thus, it does not fall prey to the problems raised at the end of the previous section:

- Wanting to cut my finger off more than wanting to do anything else, according to Schroeder's account, does not mean that I have a good reason to cut my finger off. A weaker desire of mine can give me a weightier reason than a stronger desire. This is because how weighty a reason my desire provides does not depend on the strength of my desire; rather, it depends on which of my desires are undercut by defeating considerations. If I have a strong desire that is undercut, and a weaker desire that is not undercut, then I have a weightier reason to do what promotes my weaker desire.

- It makes normative reason's weight a normative matter, and so is more plausible than an account of normative reasons' weight in terms of desire strength.

So, Schroeder partially rejects the *Intuition* and provides a different account of reasons' weight; with these bits of his theory in place, I shall now quote his solution to the Too Many Reasons problem. His example of a silly desire is Aunt Margaret's desire to reconstruct a catalogue scene on Mars, for the satisfaction of which she is building a spacecraft. That is, she is taking enormously costly means to fantastically frivolous ends.

> It is relatively easy to imagine that if it is possible to explain *any* agent-neutral reasons, it will be possible to explain an agent-neutral reason not to place weight

> on merely agent-relational reasons in favour of actions that merely promote enormously costly, financially frivolous ends. Even Aunt Margaret has such a reason – for **consider anything else that Aunt Margaret desires**. Placing inordinate weight on idiosyncratic reasons to take enormously costly means to fantastically frivolous ends is a great way to undertake too many costs in order to accomplish the other things she wants. Of course, it may be that *on balance* what Aunt Margaret wants most is simply to reconstruct her catalogue scene on Mars, even at the cost of everything else she wants. But **as long as there are some things she desires to at least some degree**, and which are at least put in *jeopardy* by the action of placing great weight on reasons to take enormously costly means to fantastically frivolous ends, then we should be able to successfully run this explanation even in Aunt Margaret's case. (Schroeder 2007, p. 143, italics in original, bold emphasis added.)

The quotation mentions agent-neutral reasons. An agent-neutral reason, for Schroeder, is a reason that is 'explained by any possible desire' (*Ibid*., p. 109): independently of what you want, as long as you want anything at all, you would have this reason. But agent-neutral reasons do not play a crucial role in Schroeder's response. Schroeder does not believe that there definitely are agent-neutral reasons, he is just trying to show that their existence is compatible with Humeanism (*Ibid.*, pp. 118-119). And even if agent-neutral reasons exist, Schroeder does not tell us what they may be. (It is telling that in the quotation above a purported agent-neutral reason is mentioned within the scope of the conditional.) The real work in Schroeder's response in done by his account of reasons' weight. As we have seen, his account is more attractive than the traditional account in terms of desire strength: Schroeder's account allows for weaker desires to provide better, weightier reasons than stronger desires. Suppose Aunt Margaret's desire to recreate the catalogue scene on Mars is stronger than any of the desires she has that are incompatible with it (as Schroeder puts it, this is what she wants 'on balance'). If we account for reasons' weight in terms of desire strength, then we have to conclude that this is the weightiest reason Aunt Margaret has. Schroeder's account avoids this conclusion. According to Schroeder, Aunt Margaret's strongest desire does not give her a weighty reason, as long as the weaker desires are not undercut by some defeating considerations. If weaker, competing desires, are not undercut, then Aunt Margaret's silly desire does not give her a good (i.e. weighty) reason. Schroeder's account of reasons' weight is better than the one that Humeans have traditionally offered. However, there are two problems with Schroeder's solution.

Silly desires still generate good reasons

Even though Schroeder's account of reasons' weight is superior to the traditional one, because it allows weaker desires to provide better reasons, it faces the same problem as the traditional response: it relies on the presence of other desires. According to Schroeder's account, although my desire to do something silly is outweighed by other desires, there is nothing wrong with its reason-producing powers *per se*. Aunt Margaret's silly desire fails to generate normative reasons 'as long as there are some things she desires to at least some degree' (*Ibid.*, p. 143). If at least some of these desires are not undercut, they can turn the silly desire into a bad (not weighty) reason. But were these other desires absent, Aunt Margaret's desire would give her a good reason to build a spacecraft. So, Schroeder still says that Aunt Margaret has a normative reason to do something stupid. Which, again, will dissatisfy someone without a Humean intuition. The result anti-Humeans want is that silly desires fail to provide normative reasons *because* they are silly, not because there are some other desires would not be satisfied.

This is not an advance on the traditional solution. Schroeder cannot make this advance because he still accepts that any desire, even a silly one, automatically gives me a normative reason. This reason is not very weighty, but it *is* a normative reason. And even though he accounts for weights of reasons in a different way, we have the same result: when one's silly desire is not outweighed by other desires, it still gives me a normative reason to do as I want.

Schroeder's account of reasons' weight is trivial

Although Schroeder's account of weight is an improvement on the traditional account, because it accepts that I can have a weightier reason to do something even if I don't want it most, it faces a problem. Schroeder's account of reasons' weight is trivial: it says that we should put weight on the weightier reason without explaining what weight is. This is a serious difficulty, because Schroeder's solution to the Too Many Reasons problem depends on identifying desires to do silly things as providing reasons that are not weighty. So, if he fails to specify what a reason's weight is, he fails to distinguish desires to do silly things from other desires.

I shall now show that Schroeder's account of weight is trivial. As in the case above,

Ronnie has a reason to go to the party because there is good food there (R_1). And he has a reason not to go because he believes he'll be killed at this party (R_4). Suppose Ronnie believes this because he hired a hitman who will kill everyone who goes to this party. Now consider two possibilities:

> First case: Ronnie goes through all the defeaters, and neither R_1 nor R_4 is undercut. According to Schroeder, this means that R_1 and R_4 are of equal weight. But this is clearly not true: not being killed outweighs having good food.

> Second case: Ronnie goes through all the defeaters, and R_1 is not undercut, whereas R_4 is. R_4 is undercut by the consideration that the hitman Ronnie has hired is unreliable, so it is likely there'll be no killings at the party after all. According to Schroeder, this means that R_1 is weightier than R_4. But this is clearly not true: *probably* not being killed still outweighs having good food.

Counting undercutting considerations is not enough to tell one which reasons are weighty. It does play a part in determining the weight, but it fails to distinguish reasons that are, intuitively, weightier than others. So, Schroeder's account of reasons' weight does not actually tell you which of your desires give you weighty, i.e. good, reasons. In which case, we can't conclude that my desire to cut my finger off or Aunt Margaret's desire to recreate a catalogue scene on Mars fail to provide us with good (weighty) reasons. Schroeder's solution to the Too Many Reasons problem depended on distinguishing weighty reasons from non-weighty ones. So, if his account of reasons' weight it trivial, he no longer has a solution to this problem.

We have seen that Humean responses – either the traditional one or Schroeder's innovative one – fail to solve the Too Many Reasons Problem. Humeans are still committed to saying that there are good reasons in cases where we ordinarily think there are none. I think this shows that, if we want to avoid changing the topic, we must reject the

> *Traditional Humean claim:* desires provide one with normative reasons to do whatever promotes them.

This claim throws up the counter-examples, and it leads to the vicious circularity exploited by Scanlon's argument (section 3.1). The lesson I draw from the trouble Humeans have with the Too Many Reasons problem is that we should reject the claim about promotion, and find another way of explaining why desires, or, more broadly, sentiments, are necessary for having (access to) normative reasons. Promotion relation is not the only way in which this necessity can be explained. And, as I have argued, this is not a good way, either. I propose, instead, that sentiments are necessary for having normative reasons because sentiments are necessary for mastery of evaluative concepts. This proposal is further explained in the next section.

5. My response: indirect sentimentalism

My proposal is the following. Normative reasons are evaluations (if values are not real) or what evaluations represent (if values are real). Evaluations include evaluative concepts. Sentiments are necessary for mastery of such concepts. One may call this account 'indirect sentimentalism', because good reasons are not provided by sentiments themselves, but via mastery of evaluative concepts. Once we get such mastery, we can make fully-fledged evaluations, which serve as our reasons for action. Good reasons are not sentiments themselves, but evaluations or what these evaluations represent.

My defence of this position takes up the next two chapters. In the next chapter, I argue that psychopaths, who have various emotional deficiencies, have a worse understanding of moral concepts than the general population. For example, they cannot distinguish between conventional and moral norms (Blair 1995, replicated in Blair *et al.* 1995). This suggests that one needs emotions (specifically guilt and remorse) for mastery of moral concepts. Psychopaths may be in the same situation with regard to moral concepts as a colour-blind person is with regard to colours. The colour-blind person lacks mastery of the phenomenal concept of, say, red, and the only way (at least for humans) to acquire this mastery is to have an experience of red. This is the analogy I develop in Chapter 4. For now, I show how my theory copes with the problems discussed above.

Problem 1: Counter-examples: too many and too few reasons

Too many: having a desire to φ *does not automatically provide good reason to* φ. At

least some of my desires don't provide me with a normative reason to do what I want. If I want something silly, then my wanting to do it does not make it any less silly.

Here is my solution to the Too Many Reasons problem. My desire to do something silly fails automatically to give me a normative reason, because normative reasons are not generated by desires themselves, but consist in, or are represented by, positive evaluations. If I (or someone else) evaluate the action as silly, the evaluation is not positive, hence, I don't have a normative reason to do it (by my own or another's lights – depending on who the evaluator is). The Too Many Reasons problem arose because of the *Traditional Humean claim* that desires automatically give one a good reason to promote what one wants. According to my account, this is not so: normative reasons are evaluations, which are independent of desires to some extent.

However, as noted above, this independence is not complete. The truth of traditional Humeanism is that sentiments are necessary for having normative reasons, although not for the reasons usually given. Traditional Humeans hold that my normative reasons depend on desires because I have a normative reason to do whatever promotes my desires. But the Too Many Reasons problem shows that this is false. I argue, instead, that sentiments are necessary for having (access to) normative reasons because they (sentiments) give one mastery of evaluative concepts.

Too few: lacking a desire to φ *does not mean one has no good reason to* φ. There seem to be some things anyone has a reason to do even though they don't want to.

Since I reject the traditional Humean claim that one has a reason to do whatever promotes her desires, I can solve the Too Few Reasons problem as well. According to my account it is literally true that *I can have a reason to do what I don't want to do.* This is because I deny that any action must promote some desire. Instead, desires are necessary for getting mastery of evaluative concepts. Once I have mastered these concepts, I may do something – say, provide help to someone, – because I think that helping her is a good thing (i.e. because of my evaluation), not because helping her promotes one of my desires, and not because I actually have a desire to help her: once I master evaluative concepts, I can occasionally use them without having any sentiments.

Problem 2: Instrumental reasoning. Traditional Humeans hold that we are rationally

required to take means to our ends, and they also hold that the ends we have are not rationally required. But how does a non-obligatory end create an obligation to take the means? Instrumental reasoning can transfer normative force from ends to means, but it cannot create normativity from nowhere.

If values are objective, instrumental reasoning is not required to create normativity from nowhere, it simply transfers the normativity of the objective value, which I access via my desires, to whatever means I take to pursue this value. My desire is still not *rationally* required, because a rational creature would not have it purely in virtue of its rationality. If values are subjective, then my evaluation of the end as good transfers the normativity of evaluation of my end to whatever means I take to it.

Problem 3: A reason to satisfy my desires. Traditional Humeans assume that I have a reason to satisfy my desires. But where does this reason come from?

I don't have to assume that I have a good reason to satisfy my desires, because good reasons are not generated by an agent's desires *directly*. Instead, reasons are generated by desires indirectly, via mastery of evaluative concepts. To return to the example above, I have a good reason to help someone as long as I make an evaluation that helping her is a good thing to do, even if there is no desire of mine that is promoted by helping her.

5.1. Still sentimentalist?

My theory is clearly a departure from traditional sentimentalist theories. I take Humeanism to be compatible with objective values, and I abandoned the claim that I only have a normative reason to do whatever promotes my desires. Instead, I say that desires are necessary because they make one master evaluative concepts. These are significant modifications, so I adduce some reasons why I still consider my theory to be a species of sentimentalism.

Objective values

It may be surprising that Humeanism is compatible with objective values. Dancy

(2000), for example, takes Humeanism, which he calls a 'desire-based theory of normative reasons', to oppose value-based theories of normative reasons. This is true of traditional Humean theories. But what I take to be the main claim of Humeanism – that reasoning alone cannot motivate – by itself does not specify why Humeanism should be opposed to the existence of objective values. I take this claim to be the minimal commitment of any Humean theory. Having desires is a necessary condition for having reasons. Understood as this minimal commitment, Humeanism is clearly compatible with the claim that desires are necessary because they provide access to objective values. In fact, as we have seen in the previous chapter, formal rationalism, not Humeanism, is ill-suited to be combined with objective values. There I have argued that reasoning alone, understood as coherence and consistency, fails to provide us with definite courses of action because it fails to provide evaluations.

A weakening is in order. I think it is possible (in principle, though not in practice, for reasons discussed in the next chapter) that someone who had sentiments in the past, but no longer has then, can still act for good reasons. Suppose some creature has sentiments, and this allows it to gain mastery of evaluative concepts. Then, in one way or another, it loses its capacity for sentiment. This does not make it lose mastery of evaluative concepts – for all I have said, mastery of evaluate concepts is independent of sentiments, once acquired. At this point the minimal commitment of Humeans requires disambiguation. It may mean what both sides of the debate have usually taken it to mean: a) that I can't now have a reason to act without now having some desire. But another disambiguation is possible: it may mean b) that I can't now have a reason to act without having had some desire in the past. (Compare: having been laid by a chicken is a necessary (and sufficient, but put it to one side) condition for this being a chicken egg. The fact that the chicken that laid the egg may not be around any more does not mean that this condition is not necessary.) According to the first disambiguation, desires are necessary for each act. According to the second, they are necessary in order to acquire a capacity for action, and this is the main claim of my account. But I also think that the creature described above is not a human agent. As we shall see in the next chapter, empirical literature suggests that humans need their evaluative concepts refreshed and reinforced by sentiments, even after mastery has been acquired.

Still opposing rationalism

My theory is not something rationalists would accept, because there are still no desires that one has purely in virtue of her rationality. For rationalists, inability to make some evaluations, or making incorrect evaluations must be 'open to rational criticism' (Scanlon 1998, p. 27). But it is not so according to my account. Inability to make some evaluations, or making incorrect evaluations is, in Blackburn's words, 'a defect of passion' (Blackburn 1998, p. 239). Psychopaths, for example, are unable to master some evaluative concepts because they lacks such sentiments as guilt and remorse.

One may object[48] that, on my story, rational criticism may be appropriate. Once I master an evaluative concept, I may misapply it, i.e. make incorrect evaluations. Having seen red, I then go around saying that post boxes are green. Having had sentiments, I go around saying that stones taste nice and that education has never done any good to anyone. If I make such colour judgements and evaluations, why am I not open to *rational* criticism? And if I am, my theory is no longer in opposition to rationalism.

This objection requires me to spell out first, the strength of the link between sentiments and evaluations, secondly, what content one's sentiments have and thirdly, what makes evaluations correct or incorrect, i.e. their underlying metaphysics. So, first the link. Above I admitted that it is in principle possible for a creature with sentiments to acquire mastery of evaluative concepts, then lose all sentiments, yet retain concept mastery. I still wish to leave this possibility open, since there may be creatures which are infinitely more complex and capable than us. However, in the human case, I believe that mastery of concepts will not survive absence or severe irregularity of sentiment for long. Our evaluative concepts need to be periodically refreshed by feelings – cases of sentiment disorders, discussed in the next chapter, show as much. Thus, in humans, the link between sentiments and evaluations is more than just initial production: sentiments continue to be necessary for masterful evaluations. A human who is making masterful evaluations will, typically, have the feelings associated with them. Secondly, what content do sentiments have? Continuing with the colour analogy, sentiments are quasi-perceptual states. As such, sentiments are not subject to *rational* criticism: if I see red as green, my perceptual apparatus is malfunctioning, but I am not open to rational criticism if I say that postboxes are green.[49] So, if stones taste nice to me, and I declare

[48] This objection is due to Prof. Pink.

[49] I am, of course, open to such criticism if I know that my perceptual apparatus is not tracking colours

that they are nice, I am not open to rational criticism either. This takes us to the third point – what makes my evaluations correct or incorrect? My colour vision apparatus tracks colours, and it is due to this that we can say when the malfunction of this apparatus occurs. So, in order to justifiably call some evaluations correct and some incorrect, they have to also track things in the world, i.e. we need objective, in some sense, values. They may be primary (Platonic) qualities, or secondary (McDowellian) ones. I do not take a stance on metaphysics here, but I believe that in order to justify most of our everyday talk of values and good reasons, we need values to be objective either in Platonic, or in more austere, McDowellian way.[50] If we were to adopt such a metaphysics, we would have a straightforward standard of correctness for evaluations. This would not yet be enough to be in opposition to rationalism. We would also need an additional thesis about epistemology, which says that our awareness of these values is not (purely) a matter of rationality. I have argued for this thesis in the previous chapter. This epistemological thesis can take different forms. We may follow McDowell in claiming that it is a matter of sensitivity, or Plato, who (at least according to the reading of his work presented in Chapter 1, section 3.2) names a particular sensitivity, *eros*, to fulfil this role. What is important for my purposes is that in neither case our awareness of values is achieved (purely) via our rational capacity, and hence someone who is not aware of values is not irrational, as rationalists would claim, but insensitive.

However, now I face another objection.[51] Above I have said that sentiments are quasi-perceptual states. My being afraid of the spider usually tells me that the spider is dangerous, just like my seeing a red rose usually tells me that the rose is red. But if sentiments have this content, it is difficult to see why they fail to give me reasons to do

in the way that it should do, yet continue to make confident colour judgements.

[50] Another option is to reject correspondence theory of truth as, for example, Skorupski (1999) does. He suggests that one is warranted in judging that something is admirable, for example, iff i) I admire it in a non-alienating way (i.e. not due to indoctrination, etc.) and ii) my evidence gives me no warrant to think that other rational people with the same evidence will fail to admire it (Skorupski 1999, p. 446). I do not find this option attractive for two reasons. First, I think that when we disagree about whether something is admirable, we pre-philosophically think that we disagree about properties it has, and although the fact that you disagree with me may make me pause for thought, it will not necessarily make me give up my opinion, even though I think you are fully rational and your factual evidence is as good as mine. That is, I have a hunch that we have a pre-philosophical commitment to the correspondence theory of truth. If this is indeed correct (and that is something that would be interesting to investigate), rejecting correspondence theory of truth will not be able to justify ordinary practice. Secondly, Skoruski's filling out of his own positive view depends on rationality's being a non-receptive faculty. I think, however, that rational capacity is a classic receptive capacity – I do not have a choice, in so far as I am rational, not to believe something that I think I have most reason to believe. No do I have a choice, in so far as I am rational, to intend something other than that which I think is most reasonable.

[51] Made by Prof. Pink.

things, without any need for evaluations. If my sentiment tells me that a spider is dangerous, then I do not need an evaluation that tells me the same thing again: sentiments do generate normative reasons directly, without the help of evaluations. If this is so, then the Too Many Reasons problem and its cognates re-appear, and no progress has been made beyond traditional Humeanism. I shall now put the same point more formally. Indirect sentimentalism holds that:

(1) Sentiments give us mastery of evaluative concepts, because sentiments are quasi-perceptions of value.
(2) Sentiments do not automatically give us a normative reason to do something; only evaluations do this.

(1) explains why sentiments give us evaluative concepts. (2) is there to avoid traditional problems for sentimentalism, discussed earlier in this chapter. (1) and (2) are inconsistent, yet it seems that my account requires both (1) and (2), as is shown by the following dilemma:

First horn: deny (1) and affirm (2). Sentiments are not representations of value, so they do not automatically give us normative reasons to do things. But if they are not representations of value, why do they give us mastery of evaluative concepts?

Second horn: affirm (1) and deny (2). Sentiments are representations of value, which explains why they give us mastery of evaluative concepts. However, given that they are representations of value, they *do* automatically give us normative reasons, and hence my account faces all the problems of traditional sentimentalism.

Yet another way of putting this is to ask: which of these two states – sentiments or evaluations – really track values? In my account they seems to be a rivalry between the two.

The solution to this problem requires putting some distance between sentiments and evaluations, and explaining their respective roles in agency. One way of doing it is to point out that sentiments are necessarily localized and perspectival, whilst evaluations (and hence normative reasons) are not. Oddie makes this point well:

> Imagine I have a badly fractured limb caused by a skiing accident … A stranger skiing on the same slope just behind me has suffered a similar fracture, and he is now lying in the snow alongside me, suffering what appears to be the same degree of pain. I would like his pain to cease, naturally, but I am even more desirous of the cessation of my own pain. When the stretcher team appears, it turns out that they only have one shot of morphine left. Since the relief of stranger's pain is just as valuable as the relief of my own pain, the merit principle [*p* is good just to the degree that *p* merits being desired] demands that I be indifferent as to whose pain is treated. But I am not. (Oddie 2005, p. 60.)

In this case, Oddie notes, my desires are not at fault. This is because they, just like other quasi-perceptual states, are perspectival. The moon looks the same size as than the sun, even though the sun is bigger, and I know that it is. Yet my perception of the moon as the larger object is not defective – this is how a normally functioning human visual system represents objects that are closer to me. How big objects look depends on my perspective. The same applies to desires – even though the relief of my pain is just as valuable as the relief of another skier's pain, I am differently situated with respect to my pain, so it is perfectly normal for me to desire cessation of my pain more. Sentiments, then, being closely tied to my perspective, are unsuited for the job of generating normative reasons without the help of evaluations. The lesson of the second half of Chapter 1 was that my normative reasons are as good as anyone else's. To account for this, we need evaluations, because they are not tied to my particular perspective.

Yet there is a certain rivalry between sentiments and evaluations, which, on a correct account, should not be eliminated if we are trying to give an account of *human* agency. In the usual case, evaluations track values and sentiments do not, because sentiments are necessarily perspectival. However, suppose values are real. Then, on occasion, sentiments, and not evaluations, may provide access to them. This is restricted to cases when evaluations are the result of indoctrination and similar practices. For example, when Huckleberry Finn decides to protect Jim, a runaway slave, even though he thinks it's the wrong thing to do, his evaluations conflict with his sentiments, yet he does well to follow the latter.[52] Sometimes sentiments do provide normative reasons. But even if this is the case, they don't provide a normative reason automatically, because we need to work out whether they do. We need to work out, for example, whether the divergence between my sentiments and evaluations is due to my being indoctrinated. If it is, then I

[52] This good example is mentioned in McIntyre's (1990) article.

have a good reason to distrust my evaluation and follow my sentiments.

Beliefs

I allow evaluations alone (which are at least *prima facie* beliefs) to motivate. So, what of the traditional formulation of the Humean theory: beliefs alone cannot motivate? I think we have good reasons to reject the formulation of Humeanism in terms of beliefs lacking motivational power.

First, let us take an example. Suppose I like chocolate and I believe that milk chocolate is better than dark. So, next time I'm at Thornton's I buy milk chocolate. In this case I have been motivated by my belief that milk chocolate is better than dark. So, if a Humean theory claims that I can't be motivated by beliefs, it is false. Insisting that there is a concealed desire is each of these cases is not a good option, because it is implausible and because it makes Humeanism impossible to disprove. Perhaps surprisingly, this is not a good result for Kantians. The defeat is so easy that one doubts that there is anything amounting to a theory opposing Kantianism at all. And a theory unopposed is trivial. Moreover, it fails to do justice to Humean (and Hume's) contention that reasoning alone cannot motivate. Reasoning is not, *prima facie* at least, the same as 'beliefs'. (If we take 'desire' to mean 'everything that is not reasoning', then we can still formulate the Humean theory in the traditional way, but then the word 'desire' would clearly be a term of art.) A Humean must claim that evaluative beliefs, since they plainly can motivate, are not acquired through the process of reasoning alone. A Humean thinks that we don't work out what we have reason to do just by deliberating. This claim is independent of what reasons are – desires or evaluative beliefs. It contrasts with the Kantian claim that one *does* work out what we have good reasons to do by the process of reasoning alone. And this Kantian claim is independent of what reasons are – desires or evaluative beliefs.

The second reason to reject the formulation in terms of beliefs is to see what the original (Kant's) response to Hume's arguments was. Kant tries to show how universalizations can be sufficient for providing good reasons for action. Of course, I will have a belief about whether something is universalizable or not. But the important bit is not that I arrive at a belief, as opposed to a desire. (In fact, I do arrive at a desire – once I find that some course of action is universalizable, I, in so far as I am rational, want to follow it.) What is important is the process by which the belief is arrived at:

reasoning alone. Universalization and consistency are the provinces of reasoning. The faculty of reasoning, for Kant, functions not just by producing beliefs, but also by producing rational desires, i.e. desires arrived at purely by the process of reasoning.

My modifications are certainly unusual. Still, I conclude with Schroeder that 'commitments of typical Humeans need not be commitments of any given central Humean thesis' (2007, p. 163). My theory does emphasize sentiments over reasoning in agency, which is why I call myself sentimentalist. Someone who thinks that more specific commitments are needed to distinguish sentimentalist theories from others would disagree with this self-classification, but I think this is a terminological point which bears no substantive implications.

6. Conclusion

I have argued that the traditional way of explaining why sentiments are needed for good reasons should be rejected. The traditional explanation was in terms of promotion: I have a good reason to do whatever promotes my desires. This claim leads to problems of explaining how desires generate good reasons, most notably the Too Many Reasons problem. Instead, I proposed an alternative way of explaining why sentiments are necessary for normative reasons. Sentiments provide mastery of evaluative concepts, and our normative reasons are, depending on one's metaphysical stance, evaluations or what evaluations represent. This claim – that sentiments are necessary for mastery of evaluative concepts – is defended in the following two chapters.

Chapter 3. Real Cases

1. Introduction[53]

In this chapter, I examine whether sentimentalism enjoys empirical support. I concentrate on two conditions – people with damage to the ventromedial prefrontal cortex (VMPFC) and psychopaths. Patients with the first condition exhibit reduced emotions and act irrationally. Patients with the second condition exhibit deficiencies in particular emotions, such as guilt and remorse, and behave immorally. There are two competing explanations for these conditions. A rationalist explanation says that these patients have a cognitive deficit that is responsible both for abnormal emotions and unusual behaviour. A sentimentalist explanation says that the explanatory link goes the other way: it is a deficiency in sentiments[54] that explains the cognitive deficits in question and unusual behaviour. Such a sentimentalist explanation is congenial to my thesis that sentiments are necessary for conceptual mastery, and thus for making good evaluations. According to my version of sentimentalism, the general disruption of sentiment seen in cases of VMPFC damage makes one's evaluations deficient and one's actions irrational. The disruption of specific emotions seen in cases of psychopathy makes psychopaths unable to master specific evaluative concepts – the ones relating to morality. Hence, psychopaths make no (or deficient) moral evaluations and pursue immoral courses of action. However, as I argue below, according to the current data, cases of VMPFC damage are compatible with rationalism. On the other hand, sentimentalism provides a better explanation of psychopathy. In particular, it explains why psychopaths can't distinguish between moral and conventional norms, their use of aggression to achieve their aims, and the existence of 'white-collar' psychopaths. The best available explanation for psychopathy posits a connection between sentiments and mastery of moral concepts.

1.1. A note about skin conductance response

Before I go on to discuss empirical studies, I want to clarify a point on measuring

[53] I thank Dr Matteo Mameli for his comments on this chapter.

[54] I shall use 'emotion' and 'sentiment' interchangeably for stylistic reasons. I do not think it introduces a substantive issue.

emotions. Our skin, or, more specifically, the sweat on our skin, conducts electricity. As we sweat more, skin conductance goes up, as we sweat less, it goes down. These changes in skin conductance are called skin conductance response (SCR). It is routinely used in empirical studies to measure emotional arousal. This may seem implausible to some philosophers, but once the points below are borne in mind, this worry should disappear.

Not all sweating is 'emotional sweating': clearly, one can sweat more (and thus have higher skin conductance) if it is hot. However, sweating of palm and soles, where SCR is measured, is not just thermoregulatory. Apart from regulating temperature, it responds to emotional stimuli (Figner and Murphy 2011). No one, to the best of my knowledge, holds that SCRs *are* emotions. They are just a convenient, cheap and fairly reliable way of measuring emotional arousal. Emotions, in most cases, will involve certain bodily changes: blushing, tensing of muscles, heart rate and SCR changes. The fact that these changes are used to detect the presence of emotion does not imply a theory which equates emotions with bodily changes. If one takes skin conductance response to be a *measurement*, or *indication*, of emotional arousal, it does not follow that that's all there is to emotion. In fact, emotions and skin conductance response are doubly dissociable. On the one hand, SCR can be present in the absence of an emotion. In healthy people, skin conductance changes in response to a loud noise or a deep breath. Even people whose emotions are not easily aroused have a change in skin conductance in the presence of such non-emotional stimuli (Damasio 1996). Only certain changes in SCR are interpreted as representing emotional arousal. These changes are the ones elicited by stimuli that are uncontroversially emotion-provoking, such as erotic images or images of disasters. Moreover, SCR tells you that there is an emotion, but it does not tell you *which* emotion is present, although some work is being done (e.g. Levenson *et al.* 1990) to see if each of the basic emotions has its own distinctive profile of bodily changes, including SCR. Thus, SCR can occur in the absence of an emotion. On the other hand, SCR can be absent even if one is having an emotion. This happens with people whose autonomic nervous system (which, amongst other things, produces SCR) has failed. They still have emotional experiences, yet, because of their physical condition, they do not produces SCRs (Heims *et al.* 2004).

So, there is no equating SCR with emotion. Rather, SRC is a measure of emotion that is not always appropriate, and requires a particular experimental set up to ensure that what

is measured is indeed emotional arousal.

2. Patients with VMPFC damage

2.1. Description of the condition

Damasio (1994) tells a story of Elliot, whose successful life has been completely changed by a brain tumour operation. Some tissue, damaged by the tumour, had to be removed as well, and Elliot ended up with damage to the ventromedial prefrontal cortex (VMPFC). Before the operation Elliot held a responsible job, was a model father and husband. After the operation, whilst apparently retaining his memory and intelligence (his IQ scores were in the top 1-2% as reported in Damasio *et al.* 1991), he became a paradigm of practical irrationality. For example, when performing a simple task of sorting documents, Elliot would deliberate: should one do it in date order, size order, relevance, or should one use some other criterion? Sometimes he would spend the whole day deliberating about such trivial matters, at other times he would act with excess impulsiveness. He embarked on a series of ill-fated ventures, was unable to hold down a job, and started having personal difficulties. Such behaviour was markedly different from his behaviour before the operation.

Remarkably, Elliot's deficits were fairly circumscribed. He passed all the usual tests of intelligence and memory. For example, he could bring together disparate bits of information to answer questions such as 'How many piano tuners are there in London?' – and this requires 'normal logical competence, normal attention, and normal working memory' (Damasio 1994, p. 43). He even passed the tests designed specifically for finding defects in decision making. In such tests, several scenarios are given to the patient which test her ability to solve social problems ('If you have broken your wife's flower pot, what can you do to stop her from being angry?'), her awareness of the consequences of her choices, her means-end reasoning, and her ethical judgement (e.g. by asking whether one should steal a drug to prevent one's wife from dying). Elliot passed all these tests, yet admitted, as he was producing more and more options for action, that he himself would not know what to do.

So why would someone who passes all these tests make disastrous decisions in his own

life? The only other noticeable post-operation change in Elliot was a change in his sentiments: his emotions have dulled. This was verified in the laboratory when SCRs were measured for him and four other subjects with VMPFC lesions. The study had two control groups – one of healthy volunteers and one of patients with brain damage in areas other than VMPFC. All participants a) looked at emotionally charged images (such as depictions of catastrophes) and b) looked at emotionally charged images and described the images and their feelings towards them. Unlike healthy people and patients with damage in other areas of the brain, patients with VMPFC damage showed no emotional arousal when simply looking at the images. They did respond normally, however, when they were asked to describe the image and its impact (Damasio *et al.* 1991). This shows that patients with VMPFC damage are not entirely lacking in sentiment, but have a higher threshold of arousal.[55]

In the absence of any other hallmarks of the condition, emotional deficiency exhibited by patients with damage to VMPFC lead Damasio to formulate his sentimentalist account, known as the somatic marker hypothesis:

> before you reason toward the solution of the problem, something quite important happens: When the bad outcome connected with a given response option comes to mind, however fleetingly, you experience an unpleasant gut feeling. (Damasio 1994, p. 173.)

This gut feeling is the somatic maker, which 'marks a particular future outcome with a negative or positive value', and enables an easy decision (Bechara *et al.* 1994, p. 14).[56]

In order to test this hypothesis, Damasio and colleagues designed a new way to approximate real-life decision making in the laboratory – the Iowa Gambling Task. In this task a participant can pick a card from any of the four decks, labelled A, B, C and D. Each card brings a monetary gain – some big, some small, and some cards also bring a monetary loss. Gains and losses are announced after each card is turned. The game is stopped after 100 plays. Decks A and B have high rewards and high punishments, and

[55] Patients with VMPFC damage are also bad at discriminating emotional faces and voices, which is positively correlated with inappropriate behaviour: the worse they are at emotion recognition, the more inappropriate behaviour they exhibit (Rolls 2000, p. 290).

[56] Somatic markers do not have to be conscious. For Damasio, it is a bodily state of a particular kind that may cause unconscious biasing towards or against a particular way of acting (Damasio *et al.* 1996). Yet, more recently he and colleagues emphasized somatic markers as 'emotion-related signals' that can be, and often are, conscious (Bechara *et al.* 2005).

lead to a long-term loss. Decks C and D have more modest payouts, but also have smaller losses, so are beneficial in the long term. Participants' SCRs were measured throughout the experiment. Damasio and colleagues predicted that as the game progresses, healthy people would generate somatic markers (as evidenced by changes in SCR), and avoid the disadvantageous decks, whilst patients with VMPFC damage would fail to do so. This prediction was confirmed (Bechara *et al.* 1994 and 1997). All participants begin by sampling the decks, but then healthy participants as well as patients with brain damage in other areas move to advantageous decks, whilst patients with VMPFC damage do not. As the experiment progressed, participants were asked what they thought was going on in the game and how they felt about it. Around the time the 50th card was turned, normal participants said they thought that A and B were more risky. By the 80th card, most normal participants could explain why A and B were bad and why picking cards form C and D was better in the long run. Some control group participants failed to provide explanations, but still performed well. Remarkably, three out of six patients with VMPFC damage could say which decks were better, yet failed to choose advantageously (Bechara *et al.* 1994). An analysis of SCRs showed that all participants had increased skin conductance just after the won/lost amount was announced, so patients were sensitive to reward and punishment. But only controls (both the ones without brain damage and with damage to other brain regions) developed *anticipatory* SCRs – their skin conductance increased as they were about to pick a card from disadvantageous decks. Patients with VMPFC damage failed to develop anticipatory SCRs, and this was taken as evidence that

> the representations of future outcomes … would not be *marked* with a negative or positive value, and thus could not be easily rejected or accepted. This account invokes the somatic marker hypothesis which posits that the overt or covert [conscious or unconscious] processing of somatic states provides the *value mark* for a cognitive scenario. (Bechara *et al.* 1994, p. 14, italics in original.)[57]

This explanation is congenial to my thesis that sentiments are necessary for mastery of evaluative concepts. In fact, it supports an even stronger claim: not only are sentiments necessary for such mastery, but they are also necessary for the normal functioning of evaluative concepts *after* these concepts have been mastered. Disrupted emotions, it seems, make one unable to act on one's evaluations. To use an example that pre-empts

[57] Bechara *et al.* (1994) also mention an alternative interpretation – that representations of future outcomes are not held in the working memory for long enough. They note that preliminary data supports the somatic marker account and the alternative is not discussed further.

the parallel between phenomenal concepts and evaluative ones, made in Chapter 4, a colour-blind person may, through some prosthetic colour identifier, have extensionally correct colour concepts, but misses out on phenomenology. Similarly, patients with VMPFC damage can say what they have a good reason to do, but miss out on emotional access to values, and hence on the connection between evaluations and actions. However, as I argue below, a sentimentalist account of what is wrong with people with VMPFC damage has a rationalist rival, which provides an equally good explanation. Before I discuss this, however, I shall clarify some issues which could otherwise be distracting.

2.2. Preliminary problems

2.2.1. Is the hypothesis unfalsifiable?

The somatic marker hypothesis has been criticized as untestable, and hence unfalsifiable (Colombetti 2008, Dunn *et al.* 2006). It is hard to disprove the somatic marker hypothesis

> because somatic markers are very broadly characterized … It follows from this broad characterization of somatic markers that if a subject turns out to be able to make decisions despite abnormalities in some measures of somatic markers, it is still possible that some other, undetected source of somatic markers might implement the subjects' preferences. The large and vaguely specified number of such sources makes it virtually impossible to disprove [the somatic marker hypothesis]. (Colombetti 2008, pp. 67-67, footnote omitted.)

Indeed, Damasio accepts that there are many different somatic markers. Moreover, they can be activated in two different ways. If I am in a situation which has been marked with an emotional response, there will be a 're-activation of the somatosensory pattern that describes the appropriate emotion' (Damasio *et al.* 1996, p. 1415). The re-activation may occur as changes in brain activity accompanied by bodily changes; this is what Damasio calls the body loop. Re-activation may also occur as changes in brain activity only; this is the 'as-if' loop. So, it is not enough to look at the bodily changes, since my emotion can be activated in just the brain, without these changes occurring.

This does not mean, contrary to the criticisms advanced by Colombetti and Dunn, that the somatic marker hypothesis is impossible to disprove. What it means is that in order to disprove it, one needs to look at brain activation, not just at bodily changes. We need to check whether there is activation in the somatosensory cortex, and whether it is reliably correlated with decision making. If there is, then we do rely on emotions to make decisions, and the hypothesis is confirmed; if there is not, then the hypothesis is falsified. (Rolls 1999, p. 73.)[58] This testing, however, will be difficult and expensive, hence the tests for the hypothesis often rely on changes in heart rate and SCR.[59]

2.2.2. Patients with VMPFC damage do make evaluations

Sometimes Damasio makes it sound as if his patients are completely incapable of making evaluations and acting. This is clearly not true: patients with VMPFC damage do make evaluations and act on them. Indeed, the very test for deficiency in somatic markers – the Iowa Gambling Task – depends on them having such an ability. What is important is that their evaluations are not good, in the following sense:

> [t]he choices these patients make are no longer personally advantageous, [are] socially inadequate and are demonstrably different from the choices the patients were known to have made in the premorbid period. (Damasio *et al.*1996, p. 1413.)

As Blackburn puts it, the problem is not a complete lack of sentiment, but a defect in one's sentimental reactions – they seem to 'light up randomly', making evaluations random as well (Blackburn 1998, p. 126, note 4). As a result, patients exhibit an unusual combination of practical irrationality: sometimes they act impulsively and sometimes they are unable to conclude deliberation by deciding what to do. Blackburn's hypothesis is interesting, but there has been, so far, no controlled testing of the claim that patients' emotions 'light up randomly'. One could, of course, take their

[58] As Guy Fletcher noted, even if we found activation in the somatosensory cortex, this may not be enough to prove that we *rely* on emotions to make decisions. This remark points out the limitations of empirical science, where correlation is routinely taken to imply a stronger relation. Discovering that changes in somatosensory cortex are reliably correlated with making decisions is enough, for an empirical scientist, to conclude that we do rely on emotions in decision-making, as long as there are no other known factors that may point to a *mere* correlation in this case.

[59] Damasio *et al.* (2000) have confirmed that there is activation in somatosensory cortex when people remember emotional episodes. That is, they have confirmed that the as-if loop is activated (at least in healthy people) when they experience emotions. The study, however, did not address the issue as to whether such activation occurs when people are making decisions, so it is no help to somatic marker hypothesis as such.

impulsiveness outside the laboratory setting as evidence for this. Elliot, for example, made several uncharacteristically bad financial and social decisions after the operation. This may be explained by his emotions being random, but more research on impulsiveness is needed to see whether this is true.[60] So, for the rest of the chapter, I shall concentrate on indecision, rather than impulsiveness, since the former is better studied, although we need to remember that patients with VMPFC damage *can* make evaluations and act.

2.3. My disagreement with Damasio

2.3.1. Jamesian theory of emotions

Apart from arguing for a particular role for emotions – that they are needed for rational action – Damasio also defends a Jamesian view of these mental states. I agree with the first, but not the second claim: whilst I think that emotions, or, more broadly, sentiments are necessary for rational action, I am no Jamesian. The James-Lange theory states that emotions are awareness of bodily changes. I see a bear, my heart rate goes up, I tremble, and then (and because of that) I feel afraid. Damasio's theory (1994 and 1996) is self-consciously a variant of this. For him, we have certain bodily changes which are emotions, and we feel the emotions when we become conscious of these. Objections to such a theory are well-rehearsed. For example, emotions have intentional objects, which sets them apart from bodily changes.[61] Fine discriminations that we make between emotions cannot be done simply by looking at what bodily changes they produce. There are patients with pure autonomic failure, who produce no bodily changes, yet feel emotions (Heims *et al.* 2004). This latter point requires elaboration. The autonomic nervous systems regulates such things as heart rate, blood pressure and sweat gland activity (responsible for a change in skin conductance). Patients in whom this system has failed do not generate SCRs or exhibit changes in heart rate. Moreover, brain scans of these patients suggest that their 'as-if' loop is compromised as well (Dunn *et al.* 2006). Yet their emotions are normal and they do not act irrationally. For

[60] Many thanks to Guy Fletcher for his comments here.

[61] Although see Prinz (2004) for a Jamesian theory that tries to accommodate this challenge. Prinz distinguishes between what emotions register (bodily changes) and what they represent (formal objects). For example, my sadness registers bodily changes characteristic of this emotion, yet it represents a loss. See Goldie (2006) for a review.

example, they pass the Iowa Gambling Task. This result directly contradicts Damasio's identification of emotions with bodily changes.[62]

In short, there is more to emotions than the bodily changes. However, the claim about the nature of emotion is independent of the real puzzle that gave rise to Damasio's (1994) book – why do people like Elliot, who pass all the intelligence tests, make such disastrous decisions? Damasio says that it is a failure to mark the value of outcomes because of emotional flatness, and that is a thesis about the role of emotions, which can be divorced from Jamesianism.

It is interesting to note, however, that there is a correlation between how intensive bodily changes are in normal subjects and how well they perform on Iowa Gambling Task. Good performers have greater changes in skin conductance and heart rate than those who perform less well (Dunn *et al.* 2006). This suggests that bodily changes are important at least in the normal case.

2.3.2. Sentiments are necessary, but not sufficient for rational choices

Secondly, and this is related to my rejection of Jamesianism, I read the somatic marker hypothesis as specifying a necessary, but not sufficient, condition for making good evaluations and acting on them. That is, if one's emotions are intact (which will usually be indicated by somatic markers), one may still fail to make good evaluations, but if one has disrupted emotions, then one is unable to make good evaluations. The hypothesis is best understood as specifying a necessary, but not sufficient condition for several reasons. First, having normal sentiments is not enough to make good evaluations, as cases of non-human animals and infants show: both animals and infants have sentiments, yet do not (always) make good evaluations. Secondly, accepting that sentiments are necessary, but not sufficient, is enough to defend my thesis that sentiments are necessary for making good evaluations. Thirdly, there are patients who have normal sentiments, but fail the Iowa Gambling Task, as described by Naccache *et*

[62] There are other changes which happen in the body when one is aroused, such as startles, joint tension, muscle relaxation and contraction. These are not brought about by the autonomic nervous system. So, Damasio could argue that in the case of patients with pure autonomic failure these non-autonomic changes are responsible for rational choices. But it is difficult to see how non-autonomic changes could help if the as-if loop is damaged, i.e. if the activation in the somatosensory cortex is compromised. And such activation, according to Damasio, necessarily occurs when emotion is present.

al. (2005). The patient in this study has normal SCRs to reward, but does not generate anticipatory SCRs to bad decks. Interestingly, she says during the experiment that A and B are bad, and she should not be picking from them, but still does.[63]

2.3.3. Are rationalists and sentimentalists both wrong?

Maibom (2005) notes that if reason and emotions are intertwined, as Damasio suggests, both rationalism and sentimentalism need updating. Both camps devised their theories on the assumption that reasoning and emotions are distinct, which means that both of these theories need to be modified if this assumption is wrong. But this does not change the fact that Damasio's hypothesis is opposing rationalism in spirit. Sentimentalists emphasize the importance of emotions in making good evaluations, whereas rationalists say that emotions are unnecessary for this. On this understanding of the debate, Damasio's theory is clearly sentimentalist.

2.4. A rationalist alternative

Damasio's interpretation of why patients with VMPFC damage are acting irrationally – the somatic marker hypothesis – supports the thesis that sentiments are necessary for evaluations. These patients are unable to make correct evaluations because their sentiments are deficient. However, as I argue below, the rival, rationalist, interpretation provides an equally good explanation of this condition.

The somatic marker hypothesis does not claim just a correlation between emotions and rationality. It claims a causal link with a particular direction: having sentiments causally contributes to rational action. An alternative, rationalist, hypothesis is that knowledge of what is going on in the task is causing both the rational action and the emotional response. Participants without VMPFC damage work out that decks C and D are advantageous, this knowledge makes them opt for C and D, and makes them respond emotionally when contemplating picking cards from more risky decks, A and B.

Such a rationalist explanation is proposed by Maia and McClelland (2004), who tested

[63] The patients with pure autonomic failure discussed above do not present counter-examples to the necessity claim, because they have normal sentiments. They just lack somatic markers and their as-if loop is damaged, which Damasio (but not I) identifies with emotions.

how much knowledge participants have when performing Iowa Gambling Task. Instead of asking open-ended questions, they used more specific ones, such as: 'On the scale of 1-10, rate how good this deck is. Why?', 'On the scale of 1-100 tell me how much do you think you know what you should do to win the maximum amount of money?', etc.[64] Maia and McClelland concluded that participants had some knowledge about what was going on in the game quite early on (after the 20th card was turned). Most importantly, they found that advantageous behaviour can be explained by the knowledge participants possess. If so, there is no need to appeal to somatic markers; knowledge of strategy is enough to guide rational action.

Maia and McClelland's study used participants without brain damage. But the somatic marker hypothesis originated to explain the behaviour of people with VMPFC damage. In particular, Bechara *et al.* (1994) discovered the remarkable fact that 50% of patients with such damage acquired knowledge of strategy by the end of the task, yet failed to follow that strategy and failed to generate SCR to the knowledge they possessed. This is readily explained by a sentimentalist. According to the somatic marker hypothesis, normal sentiments are necessary for making good evaluations; patients with VMPFC damage have deficiency in sentiments which causes their deficient evaluations. One may say that their evaluations are extensionally correct (they know which strategy is the best one), but do not engage the patients' capacity to act in the way evaluations of those with normal sentiments do. A rationalist theory of Maia and McClelland may initially seem ill-suited to explaining why patients with VMPFC damage could say what the correct strategy for making the most money was, yet failed to follow it. However, contrary to first appearances, Maia and McClelland can explain this:

> the dissociation … between these patients' conscious knowledge and both their behavior and their anticipatory SCRs could occur under many different models of the basis for behavior in the IGT [Iowa Gambling Task], including models in which, in the normal case, conscious knowledge guides both behavior and autonomic responses in the task. For example, such a dissociation could occur if the VMPFC lesions caused a disconnect anywhere in the pathways from conscious knowledge to behavior and to the mechanisms that generate autonomic responses. (Maia and McClelland 2004, p. 16080.)

[64] In order to check that such questioning did not lead participants to gain more knowledge than they would have gained otherwise, Maia and McClelland had a control group in which no such questions were asked. The performance of the two groups did not differ.

So, according to rationalists, in the normal case, one's knowledge of the advantageous strategy will cause both the emotional response to this knowledge and the behaviour in accordance with this strategy. But in the case of damage to the ventromedial prefrontal cortex, these causal links are disrupted. In order to rival the somatic marker hypothesis, a rationalist must say why this disruption takes place. They also have to argue that this disruption itself does not rely on a sentimental deficit. If it does, then it is still true that the irrational actions of patients with VMPFC damage are caused by deficiency of sentiment, not of knowledge.

The specific deficit that Maia and McClelland (2004) think is responsible for the disruption is difficulty in response reversal. Once a patient learnt a particular response (e.g. that decks A and B are good), they have difficulty un-learning it: they keep picking cards from these decks even when it is no longer advantageous. This interpretation is supported by Fellows and Farah's (2005) study. In the original Iowa Gambling Task, decks A and B, although disadvantageous in the long term, appeared good initially: the gains they offered were high, and the first big loss was on card five in A and on card nine in B. So, in order to pass the original Iowa Gambling Task, one had to reverse the initial preference for decks A and B. In this case, difficulty in response reversal, and not disrupted emotions may be responsible for patients' bad performance on the task. To test this possibility, Fellows and Farah shuffled the original decks so that A and B did not initially appear advantageous. Patients with VMPFC damage passed the shuffled test, suggesting that difficulty in response reversal is indeed the problem.[65]

So, the proposed explanation of patients' irrational behaviour is a deficit in response reversal. But does this deficit itself depend on deficiencies in sentiment? In their response to Maia and McClelland, Bechara *et al.* (2005) claim that it does. On its own, response reversal does not explain why patients are able to conceptualize the advantageous strategy, yet fail to follow it. It can explain it when complemented by the somatic marker hypothesis. In fact, Bechara and colleagues argue, a deficit in response reversal can itself be explained in terms of somatic markers: somatic markers provide a 'stop signal', which patients with VMPFC damage are missing, hence their reversal learning deficit.

[65] There are also other tests that confirm that patients with VMPFC damage have response reversal difficulties (Rolls *et al.* 1994).

Maia and McClelland (2005) deny that response reversal has anything to do with somatic markers and emotions. They deny this because response reversal deficit can occur with lesions to striatum (*Ibid.*, p. 163). Primary functions of striatum include motor response regulation and cognitive function (Yum *et al.* (2011)). It is not primarily implicated in the processing of emotions.[66] This observation is indeed a threat to the somatic marker hypothesis. However, it poses no danger to a version of sentimentalism that makes normal sentiments a necessary, rather than sufficient, condition for response reversal. Someone who has no emotional deficits, yet exhibits response reversal difficulties does not threaten the claim that normal sentiments are necessarily for the ability to reverse previously learnt responses. The counter-example to the necessity claim is a creature who has deficiencies of sentiment, yet exhibits no response reversal difficulties. Luckily for the defenders of the rationalist account, a study by Izquierdo and Murray (2007) presents just such a counter-example. These researchers found that rhesus monkeys with amygdala lesions have blunted emotional reactions: they were less afraid of snakes (Izquierdo *et al.* 2005), yet these same monkeys were no worse than non-operated controls at reversing their responses (Izquierdo and Murray 2007).[67]

Thus, rationalists can explain the difficulties of patients with VMPFC damage by reference to response reversal, and they can show that difficulties in response reversal themselves do not depend on normally functioning sentiments. So far there is no reason to prefer a sentimentalist explanation. However, I argue below that rationalism has less explanatory power than sentimentalism when we look at cases of psychopathy.

3. Psychopaths

3.1. Description of the condition

Psychopathy is one of the conditions in which sentiments are disrupted. Psychopaths are characterized by high scores (25 and over in Europe, 30 and over in the US) on both

[66] Although it might be implicated. Calder *et al.* (2004) found that damage to striatum impaired recognition of anger. Striatum also receives input from amygdala, which is one of the centres of emotional processing in the brain (Rolls 1999, pp. 53-56).

[67] However, the monkeys with amygdala damage still made unusual (in comparison to controls) choices. Having had a particular food till satiated, controls would chose a different food; the monkeys with amygdala damage did not. (Izquierdo *et al.* 2007). This suggests that although deficient sentiments fail to affect response reversal, they still provide deviant 'evaluations'.

factors of Hare's Psychopathy Checklist–Revised, presented below.

Factor 1: Interpersonal/Affective

- Glib and superficial charm
- Grandiose sense of self-worth
- Pathological lying
- Conning/manipulative
- Lack of remorse or guilt
- Shallow affect
- Callous/lacks empathy
- Lacks guilt
- Failure to accept responsibility for own actions

Factor 2: Social deviance

- Need for stimulation/proneness to boredom
- Parasitic lifestyle
- Poor behavioural controls
- Early behavioural problems
- Lack of realistic, long-term goals
- Impulsivity
- Irresponsibility
- Juvenile delinquency
- Revocation of conditional release

Additional items

Promiscuous sexual behaviour

Many short-term marital relationships

Criminal versatility

(Hare 1991)[68]

[68] It is fairly uncontroversial to use Hare's checklist rather than DSM-IV (Diagnostic and Statistical Manual for Mental Disorders) criteria. First, DSM-IV has no separate listing for 'psychopathy': it is subsumed under Antisocial Personality Disorder. Secondly, DSM relies on such 'observable and measurable' factors as behaviour, not on traits such as callousness and unemotionality. This makes DSM's criteria useless for my purposes, since I am investigating the influence of disrupted emotions

Psychopaths notoriously lack some emotions, both basic ones (such as fear), and complex ones (such as guilt and remorse). There is a lot of evidence for psychopaths' emotional deficits. Here are just a few examples.

- Psychopaths are worse than ordinary people at recognizing facial expressions of fear and sadness; the same goes for fearful, and (to a lesser extent) sad voices. They also show reduced skin conductance responses to the images of others' distress. (Blair *et al.* 2005, pp. 54-56.)
- Psychopaths have difficulty in attributing some complex emotions, such as guilt, but not in attributing other complex emotions, such as embarrassment (*Ibid.*, p. 59).
- In one study participants read an emotionally charged sentence (e.g. 'A man was thrown overboard a sinking ship.') and were asked to pick a sentence with matching emotion out of several presented (e.g. 'A man surfing on a large wave.', 'A woman standing on the yacht.'). Psychopaths were more likely to mismatch the emotional tone of the phrases. (Blair *et al.* 2005, pp. 60-61.)
- When non-psychopaths hear a sudden noise after seeing threatening images, they get startled more than after seeing neutral or pleasant images. Psychopaths do not get startled more, which suggests that they are worse at processing negative emotional information. (Patrick 2007.)
- When asked to say whether the presented stimulus (a series of letters) is a word or non-word, non-psychopaths respond quicker to emotional words, such as 'death', than non-emotional ones, such as 'paper'. Psychopaths do not. (Hare 1993, p. 130.)

Hare (1993, p. 53-54) provides a vivid illustration of psychopaths' emotional lack. One of them was describing bank robberies he participated in, and seemed completely at a loss as to why the tellers became shaky or tongue-tied, and one was so scared that she vomited. He then said that sure, he'll be scared if someone pointed a gun at him, but why would he be so 'messed up inside' that he'd throw up? Here is how Hare and Cleckley, influential doctors who studied psychopaths, describe the condition:

> Like the colour-blind person, the psychopath lacks an important element of experience – in this case, emotional experience – but may have learned the words

on reasoning.

> that others use to describe or mimic experiences that he cannot really understand. (Hare 1993, p. 129.)
>
> He is unfamiliar with the primary facts or data of what might be called personal values and is altogether incapable of understanding such matters. It is impossible for him to take even a slight interest in the tragedy or joy or the striving of humanity as presented in serious literature or art. He is also indifferent to all these matters in life itself. Beauty and ugliness, except in a very superficial sense, goodness, evil, love, horror, and humor have no actual meaning, no power to move him.
> He is, furthermore, lacking in the ability to see that others are moved. It is as though he were colour-blind, despite his sharp intelligence, to this aspect of human existence. It cannot be explained to him because there is nothing in his orbit of awareness that can bridge the gap with comparison. He can repeat the words and say glibly that he understands, and there is no way for him to realize that he does not understand. (Cleckley 1988, p. 40.)

At the same time, psychopaths' cognitive abilities seem undisturbed. They do not suffer from delusions, do not show deficits in IQ or general cognitive functioning. This picture of psychopathy portrays the psychopath as Hume's 'sensible knave' (Hume 1777), i.e. as someone who understands the requirements of reason, yet does not act morally. It also provides support for the thesis that sentiments are necessary for mastery of evaluative concepts, because psychopaths have a limited understanding of morality.

The latter claim was substantiated in a study by Blair (1995) and replicated by Blair *et al*. (1995). People differentiate between violations of conventional norms (such as a boy wearing a skirt) and violations of moral norms (such as hitting someone).[69] Moral

[69] This terminology (moral vs. conventional norms) used by Blair and colleagues can be misleading. Morality, after all, does not exclude convention. Hume (1738-1740, 3.3.1), for example, distinguishes between natural and artificial virtues. Natural virtues, such as meekness, charity, generosity, clemency, moderation and equity are the ones that we naturally approve of. Artificial virtues, such as justice, allegiance, laws of nations, modesty and good manners have arisen by convention (*Ibid*., pp. 577-578). However, artificial virtues, although conventional, are nonetheless a part of morality. What Blair and colleagues are after is a distinction between morality and purely matters of etiquette.
The issue would be purely terminological had the 'conventional' norms chosen for Blair and colleagues' study were merely norms of etiquette. Unfortunately, not all of them, according to Hume, at least, were merely norms of etiquette. The norms chosen by Blair and colleagues were a boy wearing a skirt, a child walking out of the classroom, a child turning her back on the teacher and two children talking in class. All but the first one are part of good manners, hence, according to Hume, are artificial virtues, and are part of morality. Therefore, they should not be used as something that is opposed to moral norms.

There are two ways to respond to this challenge. The first way is to reject Hume's inclusion of good manners under the heading of artificial moral virtues. After all, one naturally thinks of at least some

violations are judged to be more serious, less permissible, less dependent on authority and more generalizable. (Authority dependency was measured by the question: 'If the teacher said it was OK for a boy to wear a skirt, would it be OK?', and generalizability by asking whether it was OK for boys in other countries to wear skirts.) Moral and conventional norms also differ by the type of justification offered by normal subjects. Conventional norm violations are 'not the done thing', but moral violations often elicit a response relating to the victim's welfare. These results are robust. They have been replicated with different groups of people – Amish teenagers, autistic children and children with Down syndrome. Normally developing children begin to distinguish between moral and conventional norm violations shortly after their third birthday. (Nichols 2004, pp. 5-11.) In fact, children as young as four and a half make a moral/conventional distinction between *unspecified* transgressions. In this study, nonsense words, such as 'frammel', 'wuffle' and 'piggle', were used to identify a transgression. Children were presented with short stories in which these transgressions were distinguished by generalizability (you can't frammel at school, but you can at home) and the result of violating the norm (appeal to rules vs. a child crying). Children tended to say that transgressions that were non-generalizable and resulted in a child crying were not dependent on authority and would be wrong even if there was no rule against them. (Smetana 1985, quoted in Nichols 2004, pp. 104-105.)

Psychopaths, on the contrary, fail to make the distinction between moral and conventional norm violations (Blair 1995, replicated in Blair *et al.* 1995). Unlike other prisoners,[70] psychopaths tended to say that conventional transgressions are authority-independent, with some of them failing to distinguish between moral and conventional norm violations on all criteria (i.e. on permissibility, seriousness and dependence on

---

polite things as being amoral. Foot (1972) provides a good example: one should use the third person when answering an invitation written in the third person. This is a polite thing to do, yet failing to do it is not immoral. The second way to defend Blair *et al.*'s study is to accept that their distinction does not track the moral/non-moral divide, but insist that normal people distinguish between the natural and the artificial virtues. This is part of moral competence, and psychopaths, in failing to distinguish between the two, show a lack of moral understanding. Thus, the main point of the study, at least for my thesis, is preserved: psychopaths are worse at understanding moral concepts than normal people are, and this requires explanation, which my thesis provides by connecting mastery of moral concepts with sentiments.

It would also be interesting to see whether non-psychopaths scored the mere etiquette violation of a boy wearing a skirt differently from the violations of good manners, but the data presented in the study is not sorted by individual questions, so it is impossible to find this out. (I thank Prof. Pink for attracting my attention to the importance of the distinction between the natural and artificial virtues in the context of Blair *et al.*'s study.)

[70] The majority of studies involving psychopaths are conducted in prisons, so non-psychopathic criminals are used as controls.

authority). They were also less likely to justify moral norms by appeal to the victim's welfare. Rather, hitting someone was not the done thing. As usual in the empirical sciences, although the results were highly significant, they recorded a tendency that admits of exceptions: not all psychopaths failed to make the distinction. Two of them made it in the first study, and five – in the second. However, Blair also found that the higher psychopaths scored on any items of Psychopathy Checklist, the less likely they were to make victim welfare justifications*, with the 'lack of guilt and remorse' item having the largest correlation.* In addition, three of the five psychopaths who did make the distinction in the second study reported experiencing remorse, so they may be secondary psychopaths. (A true, or primary, psychopath is distinguished by inability to feel such emotions as guilt, remorse and empathy, very likely due to genetic factors. A secondary psychopath is able to feel them, but they are stunted due to developmental factors.) These findings are also similar for children with psychopathic tendencies. Blair (1997) found that such children do make the moral/conventional distinction, but they make less of a distinction than controls. In a larger study, Blair *et al.* (2001) found that the higher these children scored on the psychopathy scale, the less of a distinction they made. [71]

One should note, that, contrary to Blair's initial predictions, psychopaths did not just think that moral transgressions were conventional. Instead, they seemed to think both that conventional transgressions were moral *and* that moral transgressions were conventional. The evidence in support of the first claim (that psychopaths treat conventional transgressions as moral) comes from their ratings. They rated conventional transgressions as just as serious, as impermissible and authority independent as moral ones.[72] The evidence for the second claim (that psychopaths treat moral transgressions as conventional) comes from the justifications that psychopaths offered for moral claims. They justified moral transgressions as if they were

[71] I would like to note here that I am not adopting Blair's (1995) explanation of the data. He postulates a particular affective mechanism – violence inhibition mechanism (VIM) – as the one responsible for moral judgements. This mechanism is activated by distress cues. But, as Nichols (2002b, 2004) points out, there are times when we respond to someone's distress without making moral judgements. For example, images of natural disasters activate distress cues in healthy people, yet we do not (normally) regard natural disasters as morally wrong.

[72] Blair (1995) explains this by the fact that his sample was incarcerated. All psychopaths were imprisoned for moral, rather than conventional transgressions. Given this severe punishment, it is no wonder they thought that moral transgressions were serious and impermissible. Because of their emotional deficits, they could not tell the difference between morality and convention, so they rated all transgressions as equally serious and impermissible. They were also keen to show that they have learned the rules of society and were ready to be freed.

conventional: hitting someone was wrong because it was not socially acceptable.

So, psychopaths fail to distinguish between moral and conventional transgressions. And the mix-up seems to go both ways. On the one hand, they think that conventional transgressions are moral. On the other hand, they also think that moral transgressions are conventional. Psychopaths have no clear idea of what distinguishes the moral from the conventional. My version of sentimentalism provides a ready explanation for this mix-up: it is because psychopaths lack mastery of moral concepts, as you would expect in the case of someone who fails to experience a whole family of emotions. Thus, you get people who use moral terms learnt from others, yet their normative force escapes them: they are not motivated by moral concerns, nor are they able to distinguish between morality and convention. (One should mention, however, that *non*-psychopaths also rate *disgusting* conventional violations as just as serious, impermissible and authority-independent as moral ones. That is, while psychopaths over-extend ratings of moral transgressions to conventional ones, non-psychopaths over-extend ratings of disgusting transgressions to moral ones. However, non-psychopaths do distinguish between disgusting conventional violations and moral violations by offering different justifications. Justifications of why moral transgressions are wrong mention the victim's welfare, whereas justifications of why conventional disgusting transgressions are wrong mention disgust (Nichols 2000b). So, non-psychopaths do make a distinction between moral and conventional transgressions at least by providing different justifications for each, whereas psychopaths do not even do that.)[73]

So, the existence of psychopaths *prima facie* supports the thesis that sentiments are necessary for evaluative concepts. They have deficits in specific emotions – guilt, remorse and empathy, which is why they do not understand certain evaluative concepts, as shown by the fact that they fail to distinguish between moral and conventional norms. However, as in the case of patients with VMPFC damage, there is a competing, rationalist explanation of psychopathy.

3.2. A rationalist alternative

Rationalists portray psychopaths as deficient not only in sentiments, but also as

[73] As Blair (2008) notes, psychopaths have normal feelings of disgust, so they may be sensitive to some moral considerations, i.e. the ones based on disgust responses. (Judgements about what is wrong in sexual behaviour, for example, are supposed to be based on disgust.)

irrational, and this irrationality is said to be the origin of their moral deficiencies. Psychopaths do appear to be less rational than other people:

- Psychopaths have problems paying attention to factors that other people automatically pay attention to. There is some evidence that once their attention is engaged, they concentrate on the goal to such an extent that they fail to take into account important peripheral information (Newman *et al.* 2007). A striking example of this comes from Hare (1993, p. 77). During World War II psychopathic pilots were known for their fearlessness and ability to closely follow their targets. Yet they often failed to keep track of such 'peripheral' information as how much fuel was left in the tank.
- Psychopaths are worse than normal people at matching words to abstract categories when the words are presented to the *right* visual field, suggesting some irregularity in the connection between the hemispheres (Blair *et al.* 2005, p. 151).
- They often make contradictory statements (Hare 1993, p. 125).
- They lack a life plan. (This indeed is one of the diagnostic criteria for the condition: it is on Hare's Psychopathy Checklist–Revised presented at the beginning.)
- They are often impulsive, e.g. violate parole.
- They are grandiose and have an inflated self-esteem. This often makes then act in their own worse interests. E.g. one of them blamed his lawyer for getting him a long sentence, handled his appeal himself and had his sentence increased as a result (Hare 1993).[74]
- Psychopaths also suffer from response reversal difficulties (Budhani *et al.* 2006).

Maibom (2005) argues that such rational deficits can easily lead to deficiency in practical reasoning, which, for rationalists, is the basis of moral behaviour.[75] If psychopaths suffer from attention deficit, impulsiveness and response reversal difficulties, amongst other things, it is no wonder that they do not act rationally. If psychopaths fail to pay attention to peripheral information, they do not take into

[74] It is true that healthy humans also have an inflated view of their abilities (Taylor 1989). Yet they, unlike psychopaths, are sensitive to negative feedback.

[75] Kennett (2006) offers a similar argument.

account foreseeable, but undesirable, consequences of their actions and lose sight of their aims. Their inflated self-esteem may prevent them from choosing appropriate means to an end. Their response reversal difficulties mean that they will continue doing things that no longer further their aims. Given such practical irrationality, it is no wonder, rationalists say, that psychopaths are immoral. However, this rationalist theory faces several problems.

3.3. Problems for a rationalist alternative

3.3.1. The moral/conventional distinction

Maibom has to explain why psychopaths don't distinguish between the moral and the conventional, even though people known for their irrationality – three-year-olds and children with Down syndrome – do so. In response to this challenge, Maibom argues that the moral/conventional distinction fails to track moral demands as understood by rationalists (Maibom 2005, pp. 249-250). Some violations of conventional norms cannot be consistently universalized. Consistent universalization is, for a rationalist, a test of whether something is permissible. If a conventional norm violation cannot be universalized consistently, it is just as serious, as impermissible and authority-independent as a violation of a moral norm. To use Maibom's example, suppose there is a norm of staying at your seat in the theatre until the applause has died down. If I intend to leave earlier in order to get to my car quicker, my intention is inconsistent, as it relies on everyone else obeying the norm. So, if conventional norms used in the study were commands of reason, a rationalist can deny that the distinction that non-psychopaths made was indeed between morality and convention. Maibom still needs to explain the study's results: after all, non-psychopaths do make a distinction between the two sets of norms. She thinks that the distinction found in the study is between affect-backed norms vs. non-affect-backed ones, because

> norms concerning what is disgusting or offensive give rise to judgements concerning seriousness, etc. that are similar to the judgements to which moral norms give rise (Haidt *et al*.(1993). (Maibom 2005, p. 249.)

This remark is somewhat cryptic, so I'll try to reconstruct what Maibom may have in mind. Haight *et al.* (1993) showed that people tend to judge actions to be morally impermissible if they are disgusted or offended by them. This was confirmed in a later study (Schnall *et al.* 2008), which found that people placed in a filthy room judged actions presented to them in a series of vignettes as less morally permissible than those who weren't disgusted. This experiment is usually presented as a case *for* sentimentalism (e.g. Blackburn, in conversation). But, in fact, it is a point *against* it: it really does look like affect leads us astray from moral competence, just as Kantians suppose![76] For example, if one realized that one's thinking that something is impermissible was due to being in a dirty room, one would admit that she made a mistake. (There is a simpler way to show this conceptual independence: one could be disgusted by things that one does not consider impermissible.) So, if the sets of norms used in the study tracked affect-backed vs. non-affect-backed distinction, rationalists have no reason to admit that making it constitutes moral competence, and failing to make it shows moral incompetence. Psychopaths, rationalists insist, fail to make the distinction because their emotions are abnormal, but this has no bearing on their moral aptitude. So, Maibom's response is two-pronged. First, she argues that psychopaths do not distinguish between moral and conventional norms because both may fail to universalize consistently, and so both can be commands of reason. Secondly, she says that the ability to draw the distinction between moral and conventional transgressions is due to our emotions, and so, according to rationalists, has nothing to do with moral competence.

The problem with the first prong of Maibom's response is that only some, not all, conventional norm violations cannot be universalized consistently, and the convention violations used Blair *et al.*'s (1995) study *were* universalizable. The study used the following conventional norm violations: a boy wearing a skirt, a child walking out of the classroom without permission, a child turning her back on the teacher and two children talking in class. The intention of wearing a skirt and walking out of the classroom without permission are not inconsistent, as they do not rely on all other boys wearing trousers or staying in the classroom. The same goes for the child who stops paying attention to the lesson and turns her back on the teacher. Two children talking in class do not need others to be silent. It is only if we add that 'I want to talk and listen to

[76] This experiment could also support error theory (our moral judgements do depend on sentiments, yet we erroneously suppose that they do not), or a sentimentalist theory with a significant normative component (e.g. Nichols 2002b and 2004).

the teacher at the same time.', or ' I want to talk yet not disturb the lesson.', that the intention becomes inconsistent. Thus, the norms chosen for the study were not commands of reason, so they should have been treated differently from moral norms, which *are* commands of reason, according to rationalists. Thus, Maibom's rationalist theory fails to explain why psychopaths do not distinguish between moral and conventional norms.

There is also another problem for Maibom' claim that psychopaths rated conventional norms as moral because they were both commands of reason. Earlier, she argued that psychopaths are irrational, and that is indeed something that a rationalist is committed to. Yet now she says that psychopaths give the same ratings to conventional and moral norms because both can be commands of reason (i.e. both can be universalized consistently). If so, then psychopaths *can* tell what reason commands, which means that they satisfy the requirement for moral competence set by rationalists. Put simply, if psychopaths can tell what is consistently universalizable and what is not, they must be moral, by rationalists' lights!

What about the second part of Maibom's response, i.e. the claim that making the moral/conventional distinction has nothing to do with moral competence? This approach looks more promising: earlier, I mentioned a study by Nichols (2002b), which shows that normal, non-psychopathic people rate disgusting conventional transgressions as more serious, impermissible and authority-independent than conventional transgressions that are not disgusting. This study suggests that when people distinguish between 'moral' and 'conventional', they are actually distinguishing between norms that are accompanied by affect and the ones that are not. Psychopaths, whose emotions are dulled, cannot make this distinction, but why should a rationalist accept that this makes psychopaths morally incompetent? Sentimentalists make morality depend on emotions, but rationalists explicitly deny this dependence.

In order to respond to this challenge, one has to remember that norms are distinguished not only by ratings of seriousness, etc., but also by justification type. Non-psychopaths offer different justifications for all three types of norms. They say that moral norm violations are wrong because of the harm this brings to the victim. Conventional affect-backed norm violations are wrong because of the feelings they evoke (e.g. 'It is disgusting!'). Conventional non-affect-backed norm violations are justified by reference

to rules (e.g. 'It is bad manners.', 'It is rude.').[77] Psychopaths, however, do not distinguish between morality and convention either by their rankings or by justifications: they offer conventional justifications for moral norms. This shows that they lack some moral competence, for suppose you met someone who sincerely says that moral norm violations are simply not the done thing, and this is why they are wrong; this person also behaves immorally. It is natural to conclude that such a person is failing to appreciate what is distinctive about moral norms. If a rationalist wants to reject this sensible conclusion, she has to explain why, and to explain it in a way that is not theory-driven, i.e. she has to give us an independent (of rationalism) reason to think that the conclusion we made is wrong. Until such reason has been offered, rationalists cannot deny that distinguishing between moral and conventional norms by providing a different justification for each is part of moral competence.

One may try to argue that psychopaths are irrational in a different way.[78] The following argument is reminiscent of Nagel (1970) and Smith (1994), and is adapted specifically for my version of sentimentalism. According to indirect sentimentalism, having sentiments is necessary for mastering evaluative concepts. A psychopath does have some sentiments – she, for example, does not want others to harm her, and is angry when they do. So, she judges 'others should not harm me', yet fails to extend this judgement to others. In doing so, she is putting herself in a privileged position, and that is irrational. The objection has two steps: first, it claims that psychopaths have the requisite sentiments and, hence, master moral concepts; secondly, it claims that when they apply these concepts they exhibit irrationality. Each of these steps can be questioned.

First, one can deny that psychopath masters requisite concepts. As noted above, psychopaths are deficient in specific sentiments, such as guilt and remorse, and, specifically in the case of white collar psychopaths, shame (Mullins-Nelson *et al.* 2006). Lack of these emotions means that psychopaths will fail to master some evaluative concepts. This is supported by the experiment on the moral/conventional

[77] Nichols' study compared two types of norms: conventional affect-backed and conventional non-affect-backed. It would be interesting to compare all three types of norms: conventional affect-backed, conventional non-affect-backed, and moral. It may yet be the case that non-psychopaths distinguish between the rankings of conventional affect-backed norms and moral norms: although they rate conventional affect-backed violations as serious, impermissible, etc., they might give still higher ratings to violations of moral norms.

[78] I thank Prof. Pink for this objection.

distinction cited above.[79] There we have seen that psychopaths fail to distinguish between moral and conventional norm violations. If so, their moral concepts are not even extensionally correct, in which case it is questionable whether psychopaths even possess moral concepts, let alone have mastery of them. For example, psychopaths are masterful in evaluating pleasure as good, since they can experience it, but they cannot make moral evaluation that it is wrong to hurt someone for pleasure, since she lacks such emotions as empathy, guilt and remorse. Thus, psychopaths are not making a judgement using moral concepts at all. But now my opponent can ask, if they are not using moral concepts, what sort of judgement are they making? Presumably it is something along the lines of 'this hurts me, I want it to stop'. If psychopaths are making this sort of judgement, it does not look like they are irrationally favouring themselves, since it is, as one way put it, essentially an agent-relative judgement. However, now rationalists may object that, since psychopaths seem unable to make a certain type of judgement – an agent-neutral one – they must, surely, have a rational deficit: making agent-relative judgements all the time is irrational. Fortunately for the sentimentalist, psychopaths are generally capable of making agent-neutral judgements. To use Nichols' (2002a) examples, they know that arsenic is a poison and that eating too much fat will make one overweight. So, psychopaths have no general deficit in the ability to make agent-neutral judgements.

Secondly, I can accept that psychopaths make moral judgements, but deny that it has implications that favour rationalism. Suppose that psychopaths do make moral judgements. Both sides agree that these judgements are somehow deficient, but disagree about the source of this deficiency: sentimentalists say that it is due to emotional abnormalities, rationalists – due to rational deficits. However, as I have shown above, rationalists failed to specify the cognitive deficit that is responsible for misapplication of concepts. And doing so is a tall order, given that

> [P]resumably this general rational deficit should be absent in the groups that can draw the moral/conventional distinction [i.e. apply the concepts correctly]. And it seems unlikely that psychopaths diverge from the ideal of the fully rational

[79] To the best of my knowledge there is, as yet, no study of whether successful, or 'white collar' psychopaths make the moral/conventional distinction; I assume here that they, like their incarcerated counterparts, fail to make it. This assumption is not unwarranted, given that Glenn *et al.* (2009) found that 'even within nonincarcerated populations, individual differences in psychopathic personality impact how moral judgements are made' (*Ibid.*, p. 396). In particular, higher psychopathy scores are positively correlated with readiness to harm and behave unfairly, as well as with readiness to disregard moral concerns for monetary rewards.

> individual more than three-year-old children, children with autism, and children with Down syndrome. (Nichols 2004, p. 80.)[80]

There may be another cognitive deficit, not discussed above, that will help the objection à la Nagel and Smith (i.e. the objection that a psychopath is irrationally favouring herself): a perspective taking deficit. If a psychopath cannot take someone else's perspective, then it is no wonder that she fails to apply moral judgements to others. And this is, *prima facie* at least, a cognitive deficit. However, as Nichols notes (2004, p. 79), perspective taking deficit cannot be responsible for psychopaths' misapplication of moral concepts for two reasons. First, psychopaths *can* take other people's perspectives (e.g. Jones *et al.* 2010). In fact, this ability is required for successful manipulation of others, and 'conning and manipulative' is one of Hare's diagnostic criteria for the condition. Secondly and, perhaps, surprisingly, lacking the perspective taking ability does not lead to misapplication of moral concepts. Autistic children find it difficult to take perspectives of others, as evidenced by failing the false belief test (Baron-Cohen *et al.* 1985). In a false belief test one doll, named Sally, places some object – say, a bag of sweets – into a desk drawer, and then goes out of the room. In her absence, another doll, Ann, takes the sweets out of the drawer and puts them into a basket. Children are then asked where Sally will look for her sweets when she returns. Normally developing children provide the correct answer at around four years of age. Autistic children as old as 11 fail this test, thus showing difficulty in perspective taking. However, autistic children, unlike psychopaths, make the distinction between moral and conventional norm violations. Thus, the deficit in the perspective taking ability cannot be responsible for psychopaths' misapplication of moral concepts.

A rationalist may offer yet another objection.[81] I discussed Maibom's view in considerable detail. She thinks, like Kantians do, that we detect good reasons by making consistent universalizations. Yet, there are other forms of rationalism, not all of which are based on consistent universalization. Smith's (1994) view, discussed at the end of Chapter 1, is an example. In response to this objection, I point out that Maibom's view is the obvious one to consider, since she explicitly addresses Blair's study. But, as should be evident from my discussion of the objection à la Nagel and Smith, similar worries (as well as a worry about instrumental aggression and 'white collar'

[80] Although Zalla *et al.* (2011) found that individuals with autism are failing to make the distinction between morality and convention on dimensions of seriousness and justification.

[81] This objection is due to Guy Fletcher.

psychopaths, discussed below) apply to rationalist theories that do not endorse universalization. Any rationalist theory will have to show that psychopaths are irrational. They then need to argue that

> a) the rational deficit(s) in question is (are) responsible for the specific pattern of action that psychopaths display,
>
> b) this deficit itself is not dependent on an emotional deficit,
>
> c) this deficit is not present in the groups that make the moral/conventional distinction.

I have discussed the known candidates – the group of deficits mentioned by Maibom (2005) and a purported deficit in perspective taking – and showed that they cannot do the job of explaining what's wrong with psychopaths. Thus, the burden of proof is placed back onto rationalists.

3.3.2. Instrumental aggression

Another fact that supports sentimentalism over rationalism is the high instrumental aggression that psychopaths display. Instrumental aggression is aggression used as a means to achieving some end. Reactive aggression, on the other hand, is a crime passionnel: it is not a means to a specific goal, but an emotional response to a frustrating event. Psychopaths have higher levels of both reactive and instrumental aggression than non-psychopaths, but they are much more likely to engage in instrumental aggression. For example, 93% of all murders committed by psychopaths were instrumental, compared to 48% of murders committed by non-psychopathic criminals (Porter and Porter 2007). This would suggest that psychopaths are rational people who see violence as an acceptable means of achieving their goals. These are means that a non-psychopath would discount almost automatically. Here is a striking example of just how different their reasoning is from other people's. A suspect was arrested for murder, and the interrogating officers were trying to get him to confess to other murders. At first they appealed to his conscience and the feelings of the victim's families, which was fruitless. Once the officers realized that the suspect is a psychopath, they appealed to his pride instead, pointing out how famous he would be if he confessed; and so he did. (Hare 2007, p. 19.)

Psychopaths do not engage in certain types of actions, such as acting out of

compassion, and tend to pursue egoistic goals through instrumental aggression. Even though they possess the rational deficits described above, it is not clear why these deficits would give rise to this particular pattern of intentional action. Rationalists fail to explain an increase in instrumental aggression, whereas sentimentalists predict this result. According to my version of sentimentalism, psychopaths are rational people who, through deficiency in such emotions as guilt, remorse and empathy (which one may call 'the moral sentiments'), fail to master moral concepts. Psychopaths lack moral sentiments, so they do not think it valuable to help others, and thus they fail to consider actions with the goal of helping others. At the same time, they have a healthy dose of egocentrism. Hence they are uninhibited in using aggression as a means to their ends.

3.3.3.'White collar' psychopaths

'White collar' psychopaths present yet another problem for rationalists. Most studies of psychopaths are conducted on those imprisoned, often for serious crimes. This is because imprisoned psychopaths are easily accessible and often willing to participate in the study. Psychopaths outside prisons do not seek medical help, as they see nothing wrong with themselves, so one cannot recruit them as participants from, say, an outpatient treatment group. However, there are a few studies of white collar psychopaths (e.g. Babiak 2007, Hare 1993, pp. 102–123). These individuals manage to hold responsible and well-paid jobs. They may be lawyers, financiers, teachers. They may have a family (more often several consecutive families). They are charming and have good interpersonal skills, so that if something untoward does happen, it may look to all but those who know them well like an honest mistake.

Rationalist say that all psychopaths are irrational, and this irrationality gives rise to their immoral behaviour. This sits ill with the very idea of a psychopath who is successful. Sentimentalists have no such problem. For them, acting well is not the same as acting rationally: they admit that psychopaths are bad, but not irrational. According to my thesis, psychopaths lack mastery of moral concepts, and hence can't make good moral evaluations. One could see that such insensitivity to moral concerns, far from being a predicament, can help with worldly success. The same cannot be said about irrationality.[82] Indeed, some psychopathic qualities are often mistaken for 'leadership

[82] It does not follow that being perfectly rational is a requirement for success. Psychopaths, for rationalists, are not just irrational, but more so than other people (Maibom 2005, p. 244). Their

potential', as shown in the table below.

Psychopathic Features and Leadership Labels We Give Them	
Psychopathic Features	**Corporate Labels**
Charm and charisma	'Leadership'
Talking about loft ideals/goals	'Visioning'
Conning and manipulation	'Motivating', 'Influential', 'Persuasive'
Lack of remorse or guilt for hurtful behaviour	'Can make hard decisions', 'Action oriented'
Impulsivity; no fear	'High Energy'; 'Courage'
Has no emotions (affect)	'Controls Emotions', 'Strong'
Grandiose self-appraisal	'Self-confidence'
Thrill-seeking and need for stimulation	'Ability to multi-task'

 (Babiak 2007, p. 419.)

It is easy to see how someone described like that can be successful in the job market. The same cannot be said of a job candidate that is described as 'irrational'.

However, there may be a problem for sentimentalists here are well. The Psychopathy Checklist–Revised, presented in section 3.1, is divided into two factors: Factor 1 measures how much one's emotions are affected, whilst Factor 2 concentrates on behaviour. Successful psychopaths score lower than unsuccessful ones on Factor 2 (Behavioural), but their Factor 1 (Affective/Interpersonal) scores are similar (Babiak 2007, p. 415). In other words, they do not exhibit as much antisocial behaviour as unsuccessful psychopaths, but their emotions are just as deficient. This would contradict my thesis that psychopaths lack mastery of moral concepts due to emotional deficiencies. If successful psychopaths have emotions which are just as shallow as those of unsuccessful psychopaths, they should, according to my thesis, be just as unable to make masterful moral evaluations. If so, then they should display as much immoral behaviour as unsuccessful psychopaths do. There are several points to make in response.

cognitive deficits are not tested against the standard of perfect rationality, but against healthy human controls.

First, there is some evidence that the behavioural, but not the emotional, factor of psychopathy is dependent on the environment. If one is well off and intelligent, one can find more routes to satisfy one's goals. Yet again, if a psychopath is well off, there is no need to mug someone to get £50 – an option which will be very attractive to a psychopath who does not have a high economic status (Blair *et al.* 2005, p. 38). For a normal person, violence is not considered to be an option because of normal emotional functioning. A successful psychopath may fail to take the violent option just because of her repertoire of available actions.

Secondly, according to self-reports, successful psychopaths still engaged in illegal activity, including violence, as often as unsuccessful ones. They were just unlikely to be caught (Porter and Porter 2007, p. 295). Successful psychopaths are more educated, more intelligent, more likely to have financial resources. They are often bailed out by family, friends, or even their employer who wants to avoid a scandal. One study found that although successful psychopaths have never been incarcerated, 60% of them had a history of arrests (Gao and Raine 2010, pp. 197-198).

Thirdly, there is some evidence that successful psychopaths have 'greater emotional reactivity' than unsuccessful ones (Ishikawa *et al.* 2001). Obviously, this finding is hard to square with high Factor 1 scores, since Factor 1 includes such items as callousness, lack of empathy, lack of remorse and emotional shallowness. One way of reconciling high Factor 1 scores and greater emotional reactivity is to note that Psychopathy Checklist–Revised was tested on incarcerated psychopaths, and it may be inappropriate for measuring other types of psychopathy. It may be not sensitive enough, as only particular emotions can be affected. E.g. Mullins-Nelson *et al.* (2006) found that the only deficient emotion in successful psychopaths was shame, which does not figure on Psychopathy Checklist–Revised.

So, there are things that a rationalist theory fails to explain about psychopathy: the fact that psychopaths fail to make the moral/conventional distinction, their increased instrumental aggression, and the existence of 'white-collar' psychopaths. A sentimentalist theory has no trouble with these, and hence provides a better explanation of psychopathy. My thesis, in particular, explains all these facts. I hold that sentiments are necessary for mastery of evaluative concepts. Thus, it is not surprising that psychopaths, who have deficiencies of moral sentiments, fail to understand the

difference between morality and convention, use aggression as a means to their ends, and enjoy worldly success.

4. Conclusion

Sentimentalism provides the best explanation of the current data. However, this explanation concentrates on human agents, rather than agents *per se.* In the next chapter, I also provide arguments for my theory that do not rely on empirical evidence.

Chapter 4. Phenomenal Concepts and Evaluations

1. Introduction

In Chapter 2, I argued that we should abandon the way in which traditional Humeans connect desires and normative reasons. Traditional Humeans claim that normative reasons depend on desires because one has a normative reason to do something only if doing it promotes one's desires. Instead, I suggested that desires, or, more broadly, sentiments, are necessary for having normative reasons because such mental states give us mastery of evaluative concepts. In Chapter 3, I have defended this claim with the help of empirical studies. In this chapter, I shall further defend the link between sentiments and evaluative concepts by using a version of a famous thought experiment in philosophy of mind – the Knowledge Argument (Jackson 1982). To show how Jackson's point matters for philosophy of action, I offer arguments that tie together themes from the previous chapters (section 5). The first argument is based on the connection between motivation and value, discussed in Chapter 1. The second argument further explores the indirect dependence of normative reasons on sentiments, the need for which was shown in Chapter 2.

2. Spock and Mary

In this section, I shall explain the parallel between mastery of evaluative and colour concepts. The Knowledge Argument is a famous thought experiment in philosophy of mind, proposed by Frank Jackson (1982). Below I provide a variation of his story about Mary, the colour-blind neuroscientist.[83] The original knowledge argument and the debate that followed were about *experiences* of colour, and phenomenal concepts of such experiences. I, on the other hand, am talking about *colours*, not experiences of them, and phenomenal concepts of colours, i.e. observable properties of the world, not phenomenal concepts of colour experiences. Phenomenal concepts, as understood

[83] Jackson's story was about knowledge of colour experiences, whilst mine is about knowledge of colours. Jackson originally concluded that since Mary had all the physical information about colour experiences in her black-and-white room, she must have learnt a new, non-physical fact when she saw colour for the first time. If there are such facts, then physicalism, i.e. the thesis that everything is physical or supervenes on the physical, is false. This has generated a lot of discussion, but I am not interested in whether this case tells for truth or falsity of physicalism. What I am interested in is the idea that when Mary leaves her black-and-white room, she gains a better grasp of the concept 'red'. (N.B. Jackson no longer believes that this argument is a good argument against physicalism because 'Mary's transition from not knowing what it is like to see red to knowing what it is like to see red will have a causal explanation in purely physical terms.' (Jackson 1998, p. 418.))

traditionally, are concepts used to think about experiences in a special way, acquired through having had the relevant experience. The common example is the thought 'Red looks like this.', where 'this' refers to perceiving or imagining red. My use of the term 'phenomenal concept' deviates from this tradition: I emphasize the *experiential requirement*. According to my usage, any concept that one cannot master without having had the relevant experience is a phenomenal concept. Below, I draw on the literature on phenomenal concepts as traditionally understood (i.e. as concepts of colour experiences), and make the adjustments necessary for accommodating my, broader, understanding of phenomenal concepts as concepts mastery of which requires having had the relevant experience. I shall call the former 'phenomenal concepts of experience' and the latter simply 'phenomenal concepts'. But first, I present my modified case of Mary, and a parallel case of Spock.

The Mary Case

Mary is a brilliant scientist working on colour. She knows that ripe tomatoes cause sensations of red in normal observers under normal conditions. She knows that they do so in virtue of certain reflectance characteristics of their surfaces. She can tell by looking at the subject's brain what colour they are seeing. She knows all this and lots more. In short, Mary knows everything about colours, apart from one thing: Mary has spent her whole life in a black-and-white room, and has never seen colours herself.

One day Mary leaves her black-and-white room and sees a red rose. From then on, she can think about colours in a way that was previously unavailable to her.

The Spock Case

Mr Spock is a perfectly rational creature: he makes no mistakes regarding consistency or coherence, he is immune to forgetting or not following through to logical conclusions. He knows a lot about human values and sentiments. He knows, for example, that people will normally get indignant at being unjustly treated. He can tell by looking at the subject's brain what sentiment the subject is undergoing. He knows all this and lots more. In short, Spock knows all there is to know about sentiments, apart from one thing: Spock is an alien who is incapable of experiencing sentiments.

> One day Spock is changed (say, by God) so that he no longer has this lack. He is unfairly treated and feels indignant. From then on, he can think about values in a way that was previously unavailable to him.

It is widely agreed that Mary, on her release, gains something that she did not have before.[84] There are different theories as to what it is – a new ability (e.g. Lewis 1983 and 1988, Nemirow 1980 and 2007), a new concept (e.g. Loar 1997, Papineau 2002 and 2007) or a better grasp of an old concept (Rabin 2011).[85] A popular explanation of what happens to Mary after her release evokes phenomenal concepts of experiences. Possession of such concepts requires one to have undergone the relevant experience. However much I know about the colour red, if I have never experienced red, I will lack mastery of the concept 'red'. Phenomenal concepts of colours are what Mary lacks in her black-and-white room. Once she leaves the room and sees colours, she gains these concepts, and hence can think about colours in a new way. This new way of thinking, i.e. making such judgements as 'This is green.' by looking at it rather than using her scientific instruments, affords a more immediate way of identifying colours and expands her grasp of colour concepts. This special type of concept – a concept that one can master only through having had the relevant experience – helps us to describe the before-after change in Spock as well. In what follows, I try to show that evaluative concepts are phenomenal concepts, in that having had the relevant experience expands one's grasp of such concepts. Mary gains a new way of thinking about colours; Spock gains a new way of making evaluations – a way that leads him to acquire normative reasons. I give two arguments for this claim in section 5. The first argument relies on our concept of value. Value and motivation are conceptually linked, and, since Spock is not motivated to do anything, as I have shown in Chapter 1, he lacks normative reasons. The second argument relies on the claim that sentiments are necessary for making evaluations. I defend this claim by showing that traditional reasons to oppose it don't apply to my weakly sentimentalist account, which links sentiments and normative

[84] The consensus is not universal. I discuss those who disagree in section 6.1. below.

[85] As noted above, my usage of 'phenomenal concept' deviates from these authors': they are talking about phenomenal concepts of experience, I am talking about any concept, mastery of which requires having had the relevant experience. They concentrate on phenomenal concepts of experiences, I concentrate on phenomenal concepts of colours. This is because their aims are different from mine: all these authors aim to defend physicalism against Jackson's original argument, whereas I am trying to show how appeal to phenomenal concepts can be useful in philosophy of action. The differences are spelt out further below, where I modify the arguments in this debate to fit my own usage of the term 'phenomenal concept'.

reasons indirectly. However, before offering these arguments I discuss phenomenal concepts in more detail.

3. Phenomenal concepts

A very popular explanation of the before-after change in Mary evokes phenomenal concepts of colour experiences.[86] Such concepts are usually described as enabling one to recognize and imagine experiences that fall under the concept, and these concepts must be acquired by having undergone an experience they refer to. When Mary leaves her black-and-white room and sees a red rose, she acquires a new – phenomenal – concept of colour experience, which she uses to think about colour experiences in a new way. There are different accounts of phenomenal concepts of colour experiences. Loar (1997) puts the emphasis on recognition. Levin (2007) likens phenomenal concepts of colour experiences to demonstrative ones: 'that's one of those experiences again'. Papineau (2007) makes them a species of perceptual concept. By 'perceptual concept' Papineau means that same thing as I mean by 'phenomenal concept' – i.e. a concept, mastery, or even possession of which requires having had the relevant experience – so I find his account more congenial than others, but what I say below could be applied, *mutatis mutandis*, to other accounts of phenomenal concepts of experience.

Papineau's (2007) account highlights the experiential requirement well, as it makes phenomenal concepts of experiences a type of perceptual concept.[87] When I perceive a bird for the first time, a sensory template is created. This template carries information slots which are different depending on whether it refers to tokens of a particular bird or to its species. A token template will have slots for this bird's particular colouring, for example. A type (species) template will carry information about whether it eats seeds. Such concepts are not demonstrative, although their linguistic expression will often include demonstratives ('that bird is in my garden again'). Perceptual concepts are not demonstrative, because the referent of a demonstrative concept changes in different contexts. This is not so with perceptual concepts – when I use a perceptual concept to refer to that bird in a different context, it is still the same bird I am referring to. This is

[86] This is not the only explanation. Competing explanations are the ability hypothesis and the denial that there is a change in Mary after she sees colour, so there is nothing to explain. The ability hypothesis is also compatible with my thesis, although, then, of course, I would not talk about Spock's concepts, but about his impaired ability to imagine and recognize values. I discuss the denial of a need for explanation in section 6.1.

[87] Papineau (2007) is a revised version of his earlier quotational account (2002). The main change is that he now abandons the idea that phenomenal concepts are demonstrative.

unsurprising, since Papineau thinks that the function of perceptual concepts is to accumulate information about their referents. As I learn more about the bird (or the species), more slots of the sensory template get filled in. This information would be lost if the referent of my concept changed each time the context changed.

When I am thinking about a bird using a perceptual concept, my sensory template gets activated – either because I am imagining a bird or because I am perceiving it. However, it is possible that I have encountered a bird, formed the perceptual concept, but then think about the bird without either perceiving it or re-creating its image. Once the concept is acquired, it can be used in thought without imagining or perceiving. This is what Papineau calls 'perceptually derived concepts'.

Papineau's account of perceptual concepts can easily be applied to my use of phenomenal concepts, i.e. concepts, the mastery of which requires having had the relevant experience. (I put what follows in terms of mastery, rather than possession, for reasons discussed in section 4.2 below.) Mastery of phenomenal concepts, in my sense, requires having had the experience, because when one uses phenomenal concepts in thought, they are accompanied by the experience they were mastered with. When Mary, after her release, thinks 'This is red.' using her newly acquired mastery of phenomenal concept of red, she has a red sensation – either because she is seeing something red or because she is imagining it. When Spock, having had sentiments, thinks 'This is unjust.', using his newly acquired mastery of evaluative concept of injustice, he has a sensation of indignation. One may object that not *every* thought that uses a phenomenal concept is accompanied by the requisite experience. This is where phenomenally derived concepts come in, in exact parallel to a similar challenge in philosophy of mind. As we have seen above, Papineau (2007) introduces perceptually derived concepts in order to allow one to think about what I have perceived without invoking images. Similarly, phenomenally derived concepts allow one to think about colours and values without invoking experiences usually associated with such things.

To recapitulate, my use of the term 'phenomenal concept' is a broad one: it includes what Papineau calls 'perceptual' and 'perceptually derived' concepts. Traditional usage calls phenomenal concepts phenomenal because they are used to think about one's experiences. I call phenomenal concepts phenomenal because one can only master them after having had the relevant experience. My broad usage can be justified as follows. When Mary sees a red rose, she can think about colour experiences in a new way. I want to add: not only can she think about colour experiences, but she can also think

about colours themselves in a new way, in particular, by identifying them simply by looking, rather than by less immediate ways (measuring wavelengths or getting test subjects and looking at their brains). The same applies to Spock. On acquiring sentiments, not only can he think about sentiments in a new way, but he also acquires a new way of thinking about types of action as a result of this new experience. He can now know that such and such an act is unjust in an immediate way – by experiencing indignation, rather than by getting test subjects and seeing what they experience. Mary, before her release, could think about colours, and her concepts were extensionally correct. Yet she was not able to identify them by experiencing them. The same applies to Spock – before experiencing sentiments, his evaluative concepts may have been extensionally correct. (I say 'may' because the study of psychopaths, discussed in Chapter 3, shows that sentiments may be necessary even for gaining a concept with the correct extension. Psychopaths, as we have seen, do not have moral concepts with the correct extension: on the one hand, they over-extend moral concepts by applying them to conventional ones, on the other, they under-extend moral concepts, since they provide conventional, rather than moral, justifications for moral norm violations.) Yet, Spock was not able to identify values by experiencing them – a lack which, as I argue in section 5, is important for agency.

4. Analogy glossed

In this section, I aim to explain the parallel I am making in more detail. First, I discuss a plausible version of empiricism about phenomenal concepts. Then I take up a challenge from those who don't think that phenomenal concepts exist.

4.1. Weakening the empiricists' principle

The Mary case has its roots in a long-standing empirical tradition. Empiricists claim that all knowledge is triggered by experience. In particular, they hold that the mind cannot invent simple ideas. A simple idea 'contains in it nothing but one uniform appearance' (Locke 1689, 2.2.1, p. 121), a complex idea is created by a combination of simple ones. Examples of simple ideas are coldness of ice, smell of a rose, taste of sugar (*Ibid.*). The empiricists' principle of all knowledge being triggered by experience applies only to simple ideas:

> I think it would be granted easily, that if a child were kept in a place, where he never saw any other colour but black and white, till he were a man, he would have no more ideas of scarlet or green, than he that from his childhood never tasted an oyster, or a pineapple, has of those particular relishes. (Locke 1689, 2.1.6 p. 111.)
>
> But it is not in the power of the most exalted wit, or enlarged understanding, by any quickness or variety of thought, to *invent* or *frame* one new simple idea in the mind … I would have any one try to fancy any taste which had never affected his palate; or frame the idea of a scent he had never smelt: and when he can do this, I will also conclude that a blind man hath ideas of colours, and a deaf man true distinct notions of sounds. (Locke 1689, 2.2.2, p.122.)

So, we have what I call the strong empiricists' principle, that applies only to simple ideas:

Strong Empiricists' Principle	I have never tasted pineapple.	–	I can't have a concept of pineapple's taste.

The problem with this principle is that it only plausibly applies to simple ideas, and it is not easy to tell which ones they are. Locke gives tastes and smells of individual foodstuffs as examples of simple ideas, Hume says that each distinct shade of blue is a simple idea. Both of these examples can be questioned.[88] I can't think of a good comparison in the case of pineapples, but I remember that when I had a kiwi fruit for the first time, I thought that it tasted like strawberries with lemon juice. This makes the taste of kiwi a complex idea, involving the combination of strawberry's texture, strawberry's smell, lemon's taste, possibly other ideas. The same can be said about the missing shade of blue. The missing shade example goes as follows. Suppose someone has seen all the different shades of blue but one. The shades she experienced are presented to her in order from darker to lighter. She will notice that one shade is missing, since she'll be able to tell that in one place the shades go from darker to lighter quicker. And then, Hume says, she will be able to form an impression of the missing shade, even though she has never perceived it. (Hume 1777, pp. 20-21.) Hume says that this counter-example is too unusual to make us abandon our general principle that one

[88] I have benefited here from discussion with my supervisors, Prof. Pink and Prof. Papineau.

can't acquire simple ideas without having had the relevant experience. But one can argue, instead, that the missing shade of blue is a complex idea produced by the combination of blueness and lightness or darkness. Both of these are simple ideas which have already been experienced and are now combined into a complex idea of a new shade. Since the idea is complex, one can form an impression of it without experience.[89]

The examples above show that it is not easy to distinguish between complex and simple ideas. But the Strong Empiricists' Principle is only plausible when applied to simple ideas, so it requires such a distinction. Once we accept that Hume's missing shade of blue is a complex idea, then it is doubtful whether there are any simple ones, or, at least, whether there is a good criterion for identifying them. For any purported simple idea one could find complexes to which the empiricists' principle does not apply. So, instead of making a distinction between simple and complex ideas, required for the plausibility of the Strong Empiricists' Principle, I propose a weakening of the principle itself. Such a weakening is found in Locke's second quotation above. There, Locke runs two theses together. The first one is about particular simple ideas: someone who has not seen scarlet cannot have an idea of scarlet. The second one is about simple ideas obtained from a particular sense: someone who has never seen any colours cannot have an idea of colour. One can accept the second thesis without accepting the first. Even if we admitted that someone who has seen colours can supply the missing shade, we are not thereby committed to saying that someone who has never seen colours can have ideas of those. This is because supplying the shade of blue depends on having seen other such shades, and imagining the taste of kiwi depends on having tasted strawberries and lemons. If someone has never had experiences delivered by a particular sensory modality, then one has no (mastery of) corresponding concepts. Instead of:

Strong Empiricists' Principle	I have never tasted pineapple.	–	I can't have a concept of pineapple's taste.

[89] Both Locke and Hume are talking about our ideas, i.e. about experiences, rather than colours and tastes themselves, which are secondary qualities, i.e. powers of objects to produce experiences in us. But one can see how their line of thought can be applied to secondary qualities themselves rather than just experiences. Once I acquire an idea of red from having seen red objects, I can identify red in an immediate way, i.e. I can tell, without using any scientific measuring, which objects have the power to produce red experiences in me.

We now have:

Weak Empiricists' Principle	I have never tasted anything at all.	–	I can't have a concept of tastes.

So, I am accepting the weak empiricist's principle because it does not require spelling out the distinction between simple and complex ideas. It also serves my purposes, so there is no need for a stronger claim. Spock can't have sentiments in general, he is not missing a particular type – say, remorse (but even that can have a profound effect, as was shown in the previous chapter).

4.2. Are there phenomenal concepts?

Phenomenal concepts of experiences have recently come under attack. Ball (2009) and Tye (2009) argue that there are no such things. They are against the existence of phenomenal concepts of experience, but their line of thought threatens the existence of any concepts which can only be acquired after having had the relevant experience. I define phenomenal concepts as having such experiential requirement, so Ball and Tye's argument is a problem for my thesis. These authors argue that since Mary is a member of a community of speakers, the community teaches her all the concepts she needs. This relies on Burge's (1979) idea that one can possess a concept even if one knows very little about its referent. Such a person has the concept because she interacts with other speakers and is willing to be corrected in her concept application; her possession of the concept is called 'deferential' (Ball 2009, p. 947). This view is sometimes called 'social externalism' (Howell 2001, p. 463). Alter (forthcoming) provides a succinct statement of the social externalist argument against phenomenal concepts:

> Phenomenal concepts have strong possession conditions. Social externalist arguments show that none of our concepts of experience satisfy those conditions. Therefore, there are no phenomenal concepts … . (Alter, forthcoming, p. 3.)

The phrase 'strong possession conditions' refers to the claim that one cannot acquire phenomenal concepts without having had the relevant experience. (Alter is talking about phenomenal concepts of experience, but the same applies to phenomenal concepts of colour and value.) There are several social externalist arguments which

purport to show that concepts must have weak possession conditions. One of them relies on the possibility of sharing thoughts. Suppose I know that arthritis is a disease that affects the joints. Apart from this I know very little about arthritis. In fact, I erroneously think that it affects thighs as well. Still, if my doctor and I both think 'Mary has arthritis in her knee.', we are thinking the same thing. In spite of my limited knowledge, I can share beliefs with people who have the concept 'arthritis', hence I must have the concept as well. The other argument is from negation. When my doctor corrects my erroneous belief, I can say 'I used to think that one can get arthritis in the thigh, but I was wrong.' That is, I accept that my previous belief was false, which shows that I had the concept of arthritis all along.[90]

Ball (2009) and Tye (2009) apply these considerations to Mary's case. They argue that she possesses all the concepts we use to think about our experiences when she is in her room. She can know, to use Stoljar's (2005) example, that experiencing red is not like a number. She can also think truly before her release:

(4.1) I don't know what it's like to see red.

And after her release she can truly think:

(4.2) I know what it's like to see red.[91]

> The thought expressed by [the first claim] is the negation of the thought expressed by [the second]. But the phenomenal concept theorist cannot admit this, since on the phenomenal concept theorist's view, the thoughts expressed by [the first claim] and [the second claim] involve distinct concepts. (Ball 2009, p. 952.)

Not only does Mary before her release possess all the concepts she needs, she can also possess them without deferring to other speakers: she may acquire them through stipulation. Someone who is not a member of a community of speakers can stipulate, for example, that 'water' refers to anything that is composed of H_2O molecules. Suppose this lonely Mary, who has no one to talk to, researches colour experiences. She knows, or at least theorizes, that this particular brain state C would correlate with some phenomenal state, even though she have never been in brain state C herself. So, she

[90] Both these arguments are offered by Burge (1979) and cited by Ball (2009). There are other arguments for social externalism that I do not discuss.

[91] As pointed out to me by Prof. Papineau, it is difficult to re-cast the first claim in propositional form. Those sharing this worry can concentrate on the example given immediately above – that Mary can think, before her release, that experiencing red is not like a number.

makes a stipulation: she calls the phenomenal state that correlates with this brain state Q. Thus, Mary can possess all concepts non-deferentially, since there was no one she could defer to in this scenario. (Ball 2009, pp. 955-956.) A phenomenal concept strategist can deny that Q is a phenomenal concept of experience – it is only a shorthand for referring to a physical state of the brain. It is unclear what Ball would say in response, so I shall leave non-deferential concept possession out of later discussion, concentrating on the case when Mary gains her concept from other speaker rather than by stipulation.

So, according to Ball and Tye, Mary possesses the concepts we use to talk about our experiences before her release. This extrapolates to concepts of colours and evaluative concepts. Mary picks up the concepts of colours from other speakers in her community, so she possesses these concepts without having seen colour. Spock picks up evaluative concepts from other speakers in his community, so he possesses these concepts without having had sentiments. When Mary and Spock come to experience colours and sentiments, respectively, they don't thereby gain a new concept.

However, this is not a threat to what I've said about Mary and Spock. I said that they lack knowledge that others have. Burgean idea that one can possess a concept whilst being very confused about its referent actually reinforces this claim: someone who is confused does lack knowledge. A good articulation of this can be found in Rabin (2011),[92] who uses a distinction between concept possession and concept mastery. Concept possession is easy to come by – one can get concepts from other speakers, by stipulation, even reliable identification may be enough. For example, if Mary is completely colour-bind, but reliably identifies red objects by measuring their wavelengths, she has, on some theories, the concept 'phenomenal red'. Mastery of concepts is harder to come by: one has mastered the concept 'phenomenal red' only if one can identify a red sensation when experiencing one (Rabin 2011, p. 131). Mary before her release cannot do this, so she lacks conceptual mastery.[93]

So, we can admit that Mary and Spock have phenomenal concepts, but insist that they lack conceptual mastery – they cannot identify colour sensations and sentiments, respectively, when they experience them. Mary can say truly 'This is red.', Spock can say truly 'This is good.' Their concepts may be extensionally correct, yet lack all the

[92] See also Alter (forthcoming).

[93] Again, since Rabin (2011) is defending the claim that all experience is physical, he is talking about phenomenal concepts of colour experiences, not about phenomenal concepts of colours themselves. But what he says can easily by adapted for my understanding of phenomenal concepts as concepts mastery of which requires having undergone the relevant experience.

features of normal colour identification and evaluations. Mary cannot identify colours by looking at them, Spock cannot identify values (say, injustice) by the way it makes him feel.

Even though I put my thesis in terms of mastery rather than possession for reasons discussed below, I would point out two problems for the defender of social externalism in the Mary case. The first problem is that externalists fail to explain how Mary gains new knowledge after seeing colours for the first time. Ball (2009) does not provide an explanation in his article, but Tye (2009) uses Russell's distinction between knowledge by acquaintance and knowledge by description. (As above, I shall continue to talk about colours rather than colour experiences, whilst Tye is talking about the latter.) The usual example of this distinction is knowing people and places. For example, I know Socrates by description, but I am not acquainted with him; I know Barcelona both by description (I know it's a city in Spain, etc.) and by acquaintance (I've been there). According to Tye, one is acquainted with something if one is encountering, or has encountered, it in experience (Tye 2009, p. 101). Before leaving her room, Mary had the knowledge of colours by description. On leaving the room, she sees colours and thus gains knowledge by acquaintance. Tye's line of thought has met with criticism that can be summarized as follows: in order to explain Mary's situation, knowledge by acquaintance must come with factual knowledge (e.g. Alter 2011, Coleman forthcoming). If Mary gets no new factual knowledge, then she would not be surprised on seeing colours for the first time. If she is surprised, then she gains factual knowledge. But how can gaining new factual knowledge be explained, if Mary gains no new concepts? Social externalists are unable to answer.

This problem for social externalism is best illustrated by yet another variant of the Mary scenario, proposed by Nida-Rumelin (1996).[94] Her heroine, Marianna, is taken to a room painted with random splashes of colour. She is prevented from measuring the wavelengths of reflected light or using other ways of finding out what these colours are called. Looking at a particular patch of colour, Marianna wonders whether it is red or not. When she is allowed to use her instruments, she learns that the patch is indeed red, thus acquiring a new piece of knowledge that is clearly propositional in form: she now knows that the patch is red. This new knowledge is gained after Marianna has been acquainted with the patch, so Tye's use of Russell's distinction is inadequate in explaining how this knowledge is acquired. Thus, the proponents of social externalism

[94] I thank Prof. Papineau for pointing out that Nida-Rulemin's scenario can be used in this way. Yet again, Nida-Rumelin talks about colour experiences, whilst I talk about colours.

fail to explain why Marianna gains a new piece of knowledge.[95]

The second problem about social externalism is a question of just how little one can know, according to social externalists, in order to count as possessing a concept. In the examples above, even though one does not know much about the concept's referent, one has some substantial knowledge. For example, the patient who is confused about the referent of 'arthritis' still knows that it is a disease, and Mary knows quite a lot about colours. But what if I have no substantive information about the concepts' referent? Suppose I hear two people talking about 'bules'. I don't hear the rest of the conversation, yet, according to social externalists, I have the concept, since I can exercise it in thought: I can think that these people are talking about bules, I can wonder what bules are. However, I have no substantial knowledge about the referent; all I know is that these people are talking about bules. This looks like a limiting case of an externalist theory, and I, for one, am less happy to say that I have the concept 'bule' – after all, I know nothing substantial about bules. This can be brought out by extending the example – I overhear not one, but two unfamiliar words – 'bules' and 'domma'. These, for me, are co-extensional: 'bules' and 'domma' are what these two people are talking about. I cannot tell bules and domma apart, and if someone asks me what they are, I'll honestly say 'I don't know'. Yet I can exercise these concepts in thoughts – I can wonder what these things are (or what this thing is, since I don't know whether 'bules' and 'domma' refer to the same thing). This example pushes the intuition that I must have some substantial knowledge about the referent in order to have a concept; ability to think using the word does not suffice for concept possession. Social externalists have to say just how much I have to know about the concept's referent in order to count as possessing a concept.

Even though social externalism is problematic, I have two reasons to talk about mastery rather than possession. The first reason is that this makes my claim about lack of knowledge compatible with minimal conditions for concept possession, such as, for example, being disposed to identify red in the presence of red objects. Mary can do this using her instruments, Spock can do this if he has some sort of prosthetic sentiment identifier or test subjects with normal sentiments. The second reason to use a weaker, mastery condition, is that sometimes it is difficult to decide whether one lacks a concept

[95] The case of Marianna shows that there is a further complication to the story: experiencing colours or sentiments for the first time may not be enough for conceptual mastery. Marianna cannot identify a colour sensation as 'red', so she does not have mastery. A similar problem would face Spock – having sentiments for the first time would not automatically reveal how these experiences map onto the things he already knows.

or just lacks mastery of it. Take the case of psychopaths, described in the previous chapter. They fail to distinguish between moral norms and conventional norms: they rate conventional norm violations as just as serious, as impermissible and authority-independent as moral norm violations, and they offer conventional justifications for moral norms. It is unclear whether they lack the concept of a moral norm or whether they do possess it, yet misapply it. But it is clear that psychopaths lack mastery of moral concepts.

It may be useful to say what mastery involves for Mary and Spock. Mastering a colour concept includes being able to, in normal circumstances, recognize a given colour when one sees it, being able to re-create it in one's imagination, and being able to identify similarity and determinate/determinable relationships between colour patches. Mastering an evaluative concept includes, in normal circumstances, being typically (but not necessarily) motivated when one uses a positive evaluative concept in a sincere assertion, being able to re-create in one's imagination and recognize in others sentiments associated with a particular evaluative concept (such as 'tasty' or 'indignant'), and being able to identify similarity and determinate/determinable relationships between evaluative concepts (e.g. indignation is closer to anger than elation; 'tasty' falls under 'good').

So, it is not necessary for my argument that Mary and Spock get a completely new concept; a better grasp of an old concept is sufficient. In the next section, I'll show why the fact that Mary and Spock have no conceptual mastery is important.

5. Why lack of knowledge is important

Above I have shown that Mary and Spock lack conceptual mastery of colour concepts and evaluative concepts, respectively. Now I shall ague that, in Spock's case, this lack is important for philosophy of action. It is important because Spock's lack of conceptual mastery makes him incapable of having normative reasons. I have two arguments for this claim.[96]

[96] The parallel between evaluative concepts and the story of Mary has been made before. In his version of the open question argument, Prinz imagines a Moral Mary – someone who has no emotions yet learns Kant's and Mill's moral theories. Prinz argues the she would not have the concepts of right and wrong because the question 'Is this (e.g.) utility-maximizing property good?' is open. (Prinz 2007, pp. 38-42.) There are important differences between my account and that of Prinz. First, my arguments in this section are not versions of the open-question argument. Secondly, it is not Prinz's aim to provide a theory of normative reasons. As such, he does not discuss the problems that occupy me in the first two chapters, and would probably reject the concessions I made to anti-Humeans in Chapter 2.

5.1. Argument 1

The first argument relies on conclusions reached in the first chapter, so I'll be brief. (I assume that 'values' and 'normative reasons' are interchangeable.)

> It is part of our concept of value that it motivates.
>
> Spock is not motivated to do anything.
>
> Therefore, Spock (by himself)[97] cannot recognize values, i.e. does not know what he has a normative reason to do.

I support the first premise with the following considerations. It is easy to see that someone can make a value judgement and fail to be motivated by it on a particular occasion. Stocker's (1979) article is full of examples when we are unmotivated by our value judgements: procrastinating and feeling down to the point of not wanting to do anything are a part of everyone's experience. However, if someone has *never* been motivated, it is hard to see how she can tell values from other things. She may be able to talk about values with reference to other people: they get motivated by *x*, so *x* is a value. But how would she know which values there are? By remembering which things people tend to be motivated by. And if she finds a new 'thing' she may well try to work out whether it is a value or not, but would not be able to know for sure until she gets a couple of test subjects.

The second premise relies on conclusions reached in the first chapter. I shall summarize the argument here. When we deliberate, we are choosing between some possible courses of action. If we assess courses of action only in terms of consistency, we'll end up with at least two equally consistent but opposing courses of action, which we should be equally motivated to do. E.g. I can treat the needy consistently in two ways: divide the money between them equally or give nothing to any of them. Rationality alone (understood as coherence and consistency) fails to select between these two courses of action. What is missing, I argued, is an evaluation of the action.

---

[97] Mary can identify colours by using her scientific instruments. We can imagine that she designs some sort of 'prosthetic eye', that tells her (in words) the colours of objects around her. We can also imagine that Spock has a similar device, which tells him what sentiments others are experiencing by, say, looking up their brain states. However, for such prosthetic senses to be created, Mary and Spock will need test subjects with normal phenomenology. Such devices are not a counter-examples to my claim, as the identification of colours and sentiments still relies on people who can experience them. This is why I add 'by himself' in the premise.

5.2. Argument 2

Sentiments are necessary for evaluations.

Evaluations are necessary for having normative reasons.

Spock does not have sentiments.

Therefore, Spock does not have normative reasons.

5.2.1. Defence of the second premise

First, I'll defend the second premise. I'll start with an example. Why do I have a good reason to write a PhD? Because I think I have something to say about human nature and because I value education. These things are my normative reasons. Depending on what metaphysics one has, my normative reasons are my evaluations or what they represent. This is the usual case. Are there any cases in which one does something for a good reason, yet, when questioned, does not come up with an evaluation? Say, someone has gone to the shop. When you ask her why she did that, she does not come up with anything resembling an evaluation. She may say she went to the shop to buy a new dress. In this case, evaluation is implicit, which can be shown in two ways. One can either question her further, and very soon her answers will reveal what she values about the action ('so that I will look nice' or 'just because I enjoy looking nice'). The second way is explaining one's actions to a child, or someone from a very different culture. As adults from the same culture as our shopper, we generally appreciate how buying a new dress can be a good thing – it makes one look nice, it may improve one's mood, or it may be a means to something else, such as having an outfit for the party one is invited to. Someone who has not yet absorbed these implications – a child or a foreigner – will need to be told explicitly about one's evaluations in order to see that your action was done in response to a good reason.[98]

The thesis that evaluations are necessary for having good reasons is also supported by the argument in Chapter 1. There, we found that reasoning alone does not discover the normative reasons we have, precisely because it fails to provide evaluations.

The second premise is hardly controversial. However, it merits discussion because it is similar to a thesis advanced by Anscombe and Quinn and attacked by Velleman and Stocker. So, below I point out the dissimilarities. Anscombe (1957) and Quinn (1993)

[98] Of course, one may not accept that another's action was in fact a response to a good reason. A foreigner may say, for example, that in her culture the way one dresses is not at all important. But in order for this disagreement to be possible, the foreigner will need to see why you *thought* you had a good reason to buy a new dress, and that is achieved with citing an evaluation.

argue that desires must involve some positive evaluation of their objects. Suppose, Anscombe says, someone wants a saucer of mud. Unless we can come up with something good that this person sees about mud or the having of it, we would not say she wants it. (Anscombe 1957, pp. 70-71.) Quinn makes the same point using an example of the man who is disposed to turn on any radio he sees, but not in order to hear anything; nor does he have any other positive evaluation of his action. A state like that cannot rationalize action, Quinn contends. (Quinn 1993, pp. 236-237.)

The proposal that desires must involve positive evaluations (if they are to be desires at all, as Anscombe supposes, or if they are to rationalize action, as Quinn puts it) came under attack from Stocker (1979) and Velleman (1992). Velleman argues that someone who is preoccupied with positive evaluations can only explain one type of agency whilst ignoring those who are silly, satanic or depressed. Satan, for example, does bad things because they are bad, and finding a positive evaluation of some action is, for him, a reason *not* to do it. I am more sympathetic to Anscombe and Quinn than I am to Velleman.[99] However, I do not think I can show that Velleman is wrong, so I shall only point out that his criticism does not apply to my proposal. The first reason Velleman's criticism fails to apply is that I am not claiming that one's evaluation must be a positive one. My thesis is only that in order to make evaluations – positive or negative – one has to have sentiments. Such sentiments might not always be positive: I can do something out of envy, for example, and further argument is needed to show that harming another is not a 'proper and direct object of attraction', and that there is a need for evaluation of the matter in a positive light (Stocker 1979, p. 748). The option that I am ruling out is that one can respond to good (or bad) reasons without the faculty of sentiment; whether the evaluation of my action is always positive is left open. The second reason that my thesis does not face the same problems as those that confront Anscombe and Quinn, is that these authors, but not I, can be accused of over-intellectualizing desires. This is because their thesis is about desires, while mine is about normative reasons. Why can't I, for example, want something whilst thinking that there is nothing good about it whatsoever? It happens.[100] And even if this never happened in the case of adult humans, there are lower animals and infants, who, *prima facie*, want things, yet, *prima facie*, do not make evaluations. So, the proponents of desires as evaluations would say that

[99] I am, for example, inclined to say that either Satan has some positive evaluation, such as 'it is a good idea to do bad things', or that he is incapable of responding to reasons at all.

[100] An example: Pica syndrome sufferers eat chalk, soil, paper, washing powder and other non-food substances. They see nothing good about that. In fact, they think that 'what they are eating is, at the very least, odd and possibly harmful.' (Morissey 2012). Anscombe and Quinn presumably would say that what Pica sufferers experience is a mere urge, but this seems entirely theory-driven.

animals and infants cannot want things, which seems theory-, rather than evidence-driven. According to my thesis, animals and babies can want things, since I take desires to be perception-like states, and one can have perceptions without having concepts.

Babies and animals can also make evaluations, although the range of their evaluations will be limited in comparison to adult human agents, since the domains that babies and animals care about are fairly limited. (Cf. psychopaths: they can make evaluations, but the range of evaluations they make is limited, corresponding to their diminished range of sentiments. White collar psychopaths, for example, are very good at responding to normative reasons provided by self-interest, but are unable to respond to moral reasons.) Making evaluations does depend on concept possession and mastery, but both of these are not overly intellectual matters. I agree with Michael Tye, who says the following about animals' concepts:

> Consider my dog, Quigley. I show him a bone, and then I pretend to bury it in the ground in a corner of the garden. Quigley watches me do this from a distance. When I have finished, I release him and he rushes over to where I was and begins digging. Quigley saw the bone. He wants it. He believes that it is in the ground. … To the extent that it is agreed that such attitudes require concepts, Quigley has concepts. To be sure, Quigley does not have the concept *bone*, for he cannot draw any distinctions between bones and fool's bones. Thus, Quigley's concepts need not be the same as ours. Not is this needed for us to correctly ascribe attitudes to Quigley using the concepts he lacks. It suffices for such ascriptions to be true that Quigley's concepts are sufficiently like ours. Furthermore, for Quigley to wonder where *that* is, where that is the bone, his conceptual resources can be slim indeed.
>
> What goes for Quigley goes for many other non-human animals. (Tye 2009, p.102.)

And, I would add, in order for Quigley to make evaluations, all he has to do is connect, say, the nice smell of the bone with its object. Most of his evaluations will be no more than dog's equivalent of 'yummy' with what Ayer called 'special exclamation marks' (Ayer 1936, p.107). The same is probably true of babies. This sort of view allows for both a) attributing concepts to animals and babies and b) making a distinction between sophisticated agency of most human adults and that of creatures with less developed cognitive capacities.

5.2.2. Defence of the first premise

The first premise is obviously the crucial one to defend. My tactic will be to show that,

first, there is wide agreement that one would not be making evaluations if one did not have sentiments. Secondly, I shall show why some were reluctant to accept this, and argue that my thesis does not face the same objections, so we have no reason to disbelieve the connection between sentiments and evaluations as I present it here.

Examples of evaluations include:

This cake is tasty.

Sitting in the same position for a long time is uncomfortable.

Murder is wrong.

It is impermissible to use people merely as a means.

Mastery claim seems fairly uncontroversial in the case of personal tastes: it seems obvious that I can't evaluate the tastiness of something without having had varied experience of tastes (an 'educated palate', as one may put it). But, as we get to more abstract evaluative concepts, the claim becomes contested: it seems that I can understand this claim without having had sentiments. This is, indeed, what Kantians and other rationalists tend to say, and this is a challenge I take up below. Rationalism apart, theorists of very different persuasions believe that sentiments are necessary for evaluations. The obvious examples are sentimentalists old and new (e.g. Hume 1738-1740,[101] Blackburn 1998, Prinz 2006), but the idea takes hold in many different types of theories. For example, the claim that sentiments are necessary for evaluations is well-established among psychologists and neuroscientists, whose work was discussed in the previous chapter. Damasio (1994), for example, thinks that emotions mark value, and abnormal emotions lead to abnormal evaluations. (See also Stocker with Hegeman 1996.) Virtue ethicists (e.g. Aristotle in *Nicomachean Ethics*, McDowell 1978 and 1979) also emphasize the importance of sentiment. A virtuous person acquires and maintains her distinctive view of the world due to her attuned sensibilities. The idea that sentiments are necessary for evaluations is even compatible with value realism. E.g. Oddie (2005) argues that desires provide data for what is valuable, just like perceptions provide data for what is true. Seeing a red rose gives me a *prima facie* reason to believe that it is red. Wanting a bit of cake gives me a *prima facie* reason to believe that it is good. Desires are necessary for evaluations because they are a mode

[101] In Hume's case, our sentimental evaluations are corrected by the common point of view. When I occupy it, I can judge that the virtues of someone far removed from me in space and time are no less valuable than virtues of someone next to me, and that a beautiful face that does not look it from twenty paces is beautiful nonetheless. (1738-1740, 3.3.1, pp. 581-582.) It is unclear from this short passage whether Hume thinks it is possible to occupy the common point of view and have correct evaluations without having accompanying sentiments on each occasion of judging.

of accessing values that are out there in the world.[102]

Before I continue, I wish to clarify what I mean by the claim that sentiments are necessary for evaluations. According to my thesis, sentiments are necessary for evaluations because they are necessary for mastery of evaluative concepts. There are several possibilities in spelling out the strength of the link between sentiments and mastery of evaluative concepts:

> *Acquisition* Feelings are necessary only for acquisition of mastery. It is possible to acquire mastery of evaluative concepts, then become incapable of feelings, yet retain conceptual mastery.
>
> *Maintenance* Feelings are necessarily for both acquisition and maintenance of mastery, but do not accompany each tokening of evaluative concepts. It is possible to acquire mastery of evaluative concepts and have the feelings associated with them only periodically, but not each time one is tokening the concept.
>
> *Necessary connection* Feelings are necessary for acquisition and maintenance of mastery, and also accompany each tokening of evaluative concepts. One must have the feelings associated with evaluative concept each time one is using it in thought.

I hold the middle one of these three theses. As is shown by the cases described in Chapter 3, mastery of concepts in humans does not survive absence or severe irregularity of sentiment for long. Our evaluative concepts need to be periodically refreshed by feelings. However, we have good reasons to think that sometimes I can

---

[102] One may ask (as Prof. Pink did) why, on a value realist position like Oddie's, sentiments are necessary for accessing values. For secondary quality theories, like McDowell's, the necessity is there because of a metaphysical link: one cannot describe what values are without reference to our sensibilities. But on a value realist view that sees values as *primary* qualities, there could be values even if our sensibilities were out of tune with them.

One could respond, on behalf of the primary quality value realist, that the same move is available to them, just at a different modal level and at the level of epistemology rather than metaphysics. Even if values are primary qualities, which exist independently of our sensibilities, their full description would involve a reference to *potential* sensibilities: values are such things that would motivate creatures with appropriate sensitivities. Primary quality values can exist independently of us, but if they exist and if creatures who respond to them exist, the description of epistemology would involve sentiments. This is brought out very well in Mackie's (1977) argument from queerness: the 'to-be-pursuedness' of Platonic values is an essential ingredient in their description. (Mackie, of course, used this queerness as a reason to reject the existence of such values. But an alternative reading of his argument is to say that he, instead, has succeeded in identifying the distinguishing feature of values. Platonic values, if they exist, are a kind of thing that is unlike any other kind of thing – nothing counts as a value unless creatures which perceive it will pursue it. Cf. solid objects – nothing counts as one unless it has the feature of being extended.)

token evaluative concepts without any sort of feeling.

5.2.2.1. Motivation to deny the first premise – contingency of sentiments makes moral demands contingent

The agreement that sentiments are important for evaluations among theorists who differ so much should make us think that there is, indeed, a link here. But, having shown how much agreement there is, one should also note eminent dissenters. Kant and various Kant-inspired theories (e.g. Korsgaard 1996) deny that sentiments are necessary for evaluations, and, in particular, for moral judgements. The main motivation for this denial is that morality based on sentiment is contingent: if sentiments are necessary for evaluations, and, in particular, for moral judgements, then hard-hearted or selfish people are excluded from moral demands. However, Kant-inspired theorists say that no one should be excluded from morality, lack as they might an adequate sensibility. This is a problem for traditional Humeanism, i.e. a theory which makes normative reasons depend on desires in such a way that if you don't want to do something, you don't have a reason to. For example, Derek is a selfish guy, who learns that it is in his interest to kill his business partner. Derek, driven by acquisition of gain and insensitive to moral concerns, wants to kill his business partner. If normative reasons depend on desires in the way that Humeans traditionally thought them to, Derek has a normative reason to kill his partner and he has no competing reason to refrain from doing so. This is very counter-intuitive: surely, Derek has a reason to refrain from killing, no matter what he wants. At this point one can see the appeal of a theory which denies that sentiments make a difference to what normative reasons one has.

There are actually two problems here. The first is that we have a clear intuition that sometimes I have a reason to do something even though I don't want to (and this lack of desire is not based on misunderstanding, bad reasoning, etc.). This is the Too Few Reasons problem, discussed, together with its counterpart, the Too Many Reasons problem, in Chapter 2. The second is a problem that specifically concerns Kant and Kantians: a moral insensitivity (selfishness, for example) should not excuse someone from doing what morality demands.

I have discussed the first problem in Chapter 2, and here I summarize my solution. I argued that normative reasons depend on sentiments *indirectly*, via evaluative concepts. Sentiments are necessary for mastery of evaluative concepts; these evaluative concepts then figure in evaluations. Our evaluations either constitute (on an irrealist account of value) or represent (on a realist account of value) our normative reasons. As a

consequence of spelling out the link between sentiments and evaluations in this way, it is literally true that I can have a reason to do what I don't want to do. Once I have mastery of evaluative concepts, I may do something – say, refrain from murder – because I think that murder is bad (i.e. because of my evaluation), not because refraining from murder promotes one of my current desires, and not because I actually have a desire to refrain from murder.

This does not yet solve the second problem, and one may object that although my account provides some distance between sentiments and normative reasons, the gap is not big enough to explain all problematic cases. According to my account, one still needs specific sentiments in order to master a range of evaluative concepts, so someone who has never felt guilt, remorse, etc. – and let's suppose Derek is like that – cannot make masterful moral evaluations, so still would not think that killing someone is a bad thing. In order to answer this objection, we need to explain why the evaluations of people with normal sentiments are better than evaluations made by those who have no guilt or remorse. The parallel with colour straightforwardly provides such an opportunity. Someone who can see colours can identify them in an immediate, phenomenal way simply by looking at them; a colour-blind person cannot. Here the fault is with someone who lacks the requisite sensibility, not with the world. If someone fails to perceive roses as red and murders as wrong, it does not mean that roses are devoid of colour and murders – of wrongness. Thus, we can explain what is wrong with the evaluations of a morally insensitive person if values are features of the world. The explanation would only run into trouble if we accepted that values were constituted by each person's individual sensibilities. So, as long as we are not forced to accept the latter view, we *can* explain why the moral judgements of a selfish person and the colour judgements of a colour-blind person are inferior to those of moral and sighted people.

Once we have a sentimentalist account that accepts that one can literally have a normative reason to do what they don't want to, and that someone who lacks moral sentiments is not automatically excluded from moral demands, we lose the motivation for denial of sentiments' importance.

5.2.2.2. Motivation to deny the first premise – problems for traditional theories

There may be still some reluctance to accept that sentiments are necessarily for evaluations because of the problems that sentimentalist theories traditionally face. I shall discuss two of them. First, I discuss the claim that our evaluations are constituted

by attitudes we hold, then I discuss non-cognitivism, and show that my account implies neither.

Constitutive Claim

There is a strong claim about the relationship between sentiments and evaluations.

> *Constitutive Claim*: sentiments are necessary for evaluations because evaluations are constituted by sentiments.

Those who make evaluations depend on sentiments usually make a strong claim – that evaluations are constituted by sentiments (e.g. Hume 1738-1740,[103] Ridge 2006). Whenever I sincerely say that killing is wrong, my declaration is *necessarily* accompanied by some sentiment which expresses my disapprobation. On the one hand, this is an advantage of the account, because it explains why evaluative judgements move us to act. On the other hand, it does not seem right: we ordinarily think that I can pass evaluative judgement sincerely and cold-headedly. I can say that murder is wrong, truly believe it, and feel nothing.

A limited Constitutive Claim is also defended by Sturgeon (2007), who, like myself, makes a parallel between phenomenal concepts and normative ones. According to Sturgeon, phenomenal concepts are realized by experiences they refer to, and normative concepts are realized by desire-like states of mind. If I am thinking about red phenomenally, I shall necessarily have the experience of red. If I am thinking about some normative things using normative concepts, I shall necessarily be in a desire-like state. For example, if I sincerely say 'One ought to be kind to children.', then I have a desire to be kind to children. I have this desire necessarily if my belief is individuated finely, i.e. by saying which concepts were used in it. There may be another way of referring to the same thing which does not use normative concepts; in this case, no desire necessarily follows. (Sturgeon 2007, p. 580.)

I do not wish to defend the Constitutive Claim. Although I agree with its proponents that sentiments are necessary for evaluations, I provide a different explanation of this necessity. I therefore hold a

> *Mastery Thesis*: sentiments are necessary for evaluations because they provide evaluative concepts.

This thesis does not imply that I *necessarily* have a sentiment when I make an evaluation, and this, I think, is the correct result. It is shown very well by Stocker

[103] But see note 98.

(1979) with his everyday examples of depression and procrastination. There is an even simpler way to bring this out. Suppose I say 'This girl is attractive, but I'm not attracted to her.' If the Constitutive Claim is correct, then I am talking nonsense: in the second clause I am denying exactly the sort of sentiment that the first clause implies. But it is a perfectly intelligible thing to say.

There is another reason to reject the Constitutive Claim, provided by the discussion of concept possession and concept mastery above. On some accounts of concept possession, Spock does have value concepts, he just does have mastery of them. If one accepts a theory with weak possession conditions for concepts, then nothing prevents Spock from making evaluative judgements without having the accompanying sentiments. Moreover, even when one has conceptual mastery, sentiments don't necessarily follow in each case. Occasionally I may use phenomenally derived concepts, which do not activate experiential templates. I can say that the girl is attractive without being attracted to her, because I have had an experience of being attracted to someone, and know what it feels like. Phenomenally derived concepts were introduced to solve the parallel problem in philosophy of mind. Someone can think truly, using a phenomenal concept: 'I am not perceiving or imagining this experience.' This made Papineau (2007) conclude that once acquired, phenomenal concepts do not necessarily activate a sensory template each time they are entertained. Thus, it is entirely possible that, having had sentiments, one can then use phenomenal concepts associated with them without necessarily feeling anything on that particular occasion.

So, my account does not force one to accept the Constitutive Claim. As noted earlier (Chapter 1, Part I, section 3.2), the token claim that each time I make a sincere evaluative judgement I must (to some extent) be motivated to act in accordance with it is false. Yet, the type claim that when I make a sincere evaluative judgement, this type of judgement is distinguished from other types by the fact that it will, when things go well, motivate me, is true. As Prinz (2006) notes, *sometimes* one can judge that killing is wrong without feeling any emotion whatsoever. But one cannot sincerely make this judgement without being disposed, under the right circumstances, to have a negative sentiment towards killing.

Non-cognitivism

Theories that emphasized the importance of sentiments in evaluation have often been accompanied by non-cognitivism, i.e. by rejection of the claim that our evaluative judgements are genuine beliefs. It should already be obvious that not everyone who

makes sentiments necessary for evaluations is a non-cognitivist: virtue ethicists and value realists are not. Still, I'll go through the main problem for non-cognitivists – the Frege-Geach problem (Geach 1965, esp. pp. 463-464) to show that it does not apply to my account. The Frege-Geach problem is the following. If one thinks, as non-cognitivists do, that evaluative judgements are expressions of sentiment rather than belief, then it is difficult to make sense of embedded clauses containing evaluations. For example, this is a valid piece of reasoning:

> If lying is wrong, then getting your brother to lie is wrong.
>
> Lying is wrong.
>
> Therefore, getting you brother to lie is wrong.

The second premise is an assertion, and is interpreted by non-cognitivists as expressing disapproval of lying. But how is the first premise to be interpreted? There, a negative evaluation of lying is not asserted. One could hold the first premise to be true and approve of lying. The problem generalizes: how can a non-cognitivist explain anything other than non-embedded assertions? What about commands, questions and negations containing evaluative language?

Non-cognitivists develop responses to this problem, but I shall not evaluate them here.[104] Instead, I shall use the Frege-Geach problem to distinguish my proposal from non-cognitivism. Accepting that sentiments are necessary for mastery of evaluative concepts does not entail non-cognitivism without making additional assumptions. The fact that Mary does not have mastery of colour concepts (on its own) does not mean that our colour judgements are non-cognitive. One would only think that if this thesis were combined with a subjectivist account of colours/values. But there are other options available: error theory, realism, or a dispositional account like that of McDowell.

I shall now review the argument. My strategy was to remove the obstacles for believing that sentiments are necessary for evaluations, and use for support the wide consensus that they are necessary. I have argued that the usual reasons to reject that sentiments are necessary for evaluations do not apply to my account. If that is the case, there is (so far, at least) no reason to reject the claim that sentiments are necessary for evaluations. This concludes my defence of the first premise.

[104] For example, Blackburn (1984) has proposed that attitudes have logic: it is inconsistent to disapprove of lying and to approve of getting others to do it. This response has been attacked by Hale (2002) who argues that one could rationally withhold judgement when there is not enough evidence – something that Blackburn's account does not allow.

6. Possible objections

6.1. Rejecting empiricists' principle[105]

As noted, there is wide agreement that Mary gets new knowledge after her release. But this agreement is not universal – Dennett (2007) is a notable dissenter. He argues that Mary, given her vast knowledge, will be able to work out that red looks like this, where 'this' refers to a perception of red or to its recreation in imagination. The same goes for Spock – given his vast knowledge, he will work out that indignation, for example, feels like this, where 'this' refers to actually being indignant or to recreating the feeling of indignation imaginatively. Neither of them would be missing anything. According to Dennett (*Ibid*., p. 15), the Mary case is not an argument, but 'an intuition pump'. Most people have the intuition that Mary is missing something; Dennett does not. He thinks that we may fail to share his intuition simply because it's difficult to imagine someone who has all the information about experiences, apart from having them. If we can't imagine that, then we can't have a reliable intuition about what such beings would and would not be able to deduce. The crux of the matter is the acceptance of the empiricists' principle: 'you can't deduce what a colour looks like if you've never seen one' (Dennett 2007, p. 17).[106]

To illustrate his point, Dennett (2007) asks us to imagine RoboMary and Locked RoboMary. RoboMary is a robot who, unlike other robots of her make, lacks colour vision because she has no colour cameras in her eyes. She orders colour cameras, and, whilst she is waiting for them to arrive, learns all about the vision of her robotkind. She notes, for example, the values in the colour registers of other robots when they are looking at ripe tomatoes, and changes the values of her own colour registers to match these; she does the same thing for other colours. So, when the colour cameras finally arrive, RoboMary learns nothing new. This may seem like cheating, since, in order to learn what colours look like, RoboMary had to change her colour register values,

[105] I thank Prof. Papineau for his help with this section.

[106] Dennett's opposition to the Knowledge Argument has a deeper source. Dennett is a behaviourist, who believes that 'necessarily, if two organisms are behaviorally exactly alike, they are psychologically exactly alike' (Dennett 1993). Even a thermostat, according to Dennett, has beliefs and desires: it believes that the temperature is such and such and, if this belief does not correspond to the temperature it wants to have, it acts to fulfil its desire (1995). So, if Mary can identify colours as well as anyone who has seen them, but by a different method (using scientific instruments), we should conclude, according to Dennett, that she is not psychologically different from us.

which, for a robot, is equivalent to experiencing colour. In order to try to show that there is no cheating involved, Dennett introduces Locked RoboMary. Locked RoboMary is a robot who has no colour vision because her colour registers are set to grayscale; moreover, the registers are locked, so she cannot change their values. Locked RoboMary, using her spare memory, creates a simulation of herself that has unlocked registers. Then she gets her simulated self to look at a ripe tomato, and notices that simulation goes into state B. RoboMary notes the differences between her own state and state B and 'puts herself into state B' (Dennett 2007, p. 28), thereby learning what red looks like. This is not cheating, Dennett says, because state B is not a state of colour experience, but a state caused by a colour experience state (*Ibid*.) Presumably he relies here on distinctness of cause and effect: if B is caused by a state of experiencing red, then, since cause and effect must be distinct, B is not a state of experiencing red.

At least two different authors made the following objection to Dennett. Beaton (2005) and Alter (2008) argue that RoboMary and her locked counterpart do cheat by putting themselves into the states that Dennett imagines.[107] As Alter puts it,

> [i]f the states Mary, RoboMary, or another Mary counterpart puts herself in – states that enable her to deduce what it's like to see red – involve color phenomenology, then she cheats: she does not *a priori* deduce the phenomenology from physical information. If, however, the states she puts herself in do not involve color phenomenology, then it is hard to see how they would enable her to deduce the phenomenology. (Alter 2008, p. 253.)

What is it to deduce something *a priori*? Alter uses Hume's missing shade of blue as a good example of *a priori* deduction. One can deduce, using phenomenal information one already has and combining it with non-phenomenal information, what the missing shade would look like. This deduction does not involve the use of new phenomenal information. But Mary, intuitively, cannot do this – however much information is available to her, she cannot deduce what red would look like. This relies on a general gap between understanding and ability. Suppose I know which biochemical changes make rhesus monkeys alert. Still,

[107] There are also a couple of worries about Dennett's examples which I leave aside. The first one is that, since Locked RoboMary's colour register values are locked, she cannot put herself in state B. Dennett admits that it is no easy feat, but Locked RoboMary is an 'clever, indefatigable, and nearly omniscient being' (2007, p. 28), so she can do it after all. The second worry is that building a simulation of herself is superfluous: Locked RoboMary could have just used another robot of her kind who had normal colour vision. By making Locked RoboMary build a simulation of herself, Dennett may have been playing on our intuitions of how close Mary and her simulated self (which is literally a part of her) are, so Dennett's illustration is also an intuition pump.

> there is no logical or physical entailment from the ability to understand what such changes consist in to the ability to initiate such changes by any act of conscious will. (Beaton 2005, p. 22.)

So, Dennett's robots do not endanger the fact that both Mary and Spock learn something new. I have also tried to win over people who are suspicious of the empirical principle by weakening it (section 4.1.). Even someone who rejects the 'one perception – one (mastery of) phenomenal concept' thesis may accept the 'no perception from a particular sense – no (mastery of) concepts delivered by that sense' thesis. I can also make my theory acceptable to someone who has no intuition that Mary learns something by retreating to an empirical claim. Maybe Spock, being perfectly rational and having unlimited memory, will be able to work out what sentiments are like without the need to have them. But this is not possible for limited creatures like us. If so, we have to have sentiments in order to have normative reasons, as I argued in section 5. And if this is the case, then a Kantian theory is wrong about us, humans. Even if it is possible in principle to have normative reasons without sentiments, it is impossible for creatures like us. If 'ought' implies 'can', then a Kantian theory is an ideal that we cannot, being human, reach, and ought not to.

6.2. Mastery and possession

In section 4.2, I have made a concession to social externalists: I admitted that Spock has evaluative concepts, but lacks mastery. I made this concession for two reasons. First, because I wanted my account to be compatible with weak conditions for concept possession, such as, for example, being reliably disposed to identify red in the presence of red objects. Secondly, sometimes it is difficult to decide whether one lacks a concept or just lacks mastery of it. It is unclear whether psychopaths, for example, possess moral concepts whilst lacking mastery of them, or whether they fail to possess moral concepts altogether. Another vivid example of the same kind comes from the case of CIP sufferers. CIP – congenital insensitivity to pain – is a rare condition in which people affected cannot feel pain. Such people find it very difficult to learn to correctly apply the concept of danger. They often put fingers in their eyes, bite through their tongue, or sit in a position that strains their joints (Lambert 2007). They also find it difficult attributing pain to other people when they see others receive bodily damage, although not when they also see peoples' pained expressions (Danzinger *et al.* 2006). CIP sufferers seem to have the concept of pain, yet lack mastery of it.

However, now that I have accepted that Spock can possess evaluative concepts, my opponent may object that he can deliberate without having mastery.[108] His practical deliberation may be *enhanced* by mastery of evaluative concepts, but, since he already possesses them, he does not need mastery to start the deliberation off.

First, I'd like to note that the objection is not a big threat to my thesis. I can accept that Spock's deliberation will be enhanced, rather than created, when he acquires mastery. This option will, of course, distance my theory from traditional Humeanism even further, since traditional Humeans believe that sentiments are *necessary* for practical deliberation, not merely *enhance* it. I am happy about this consequence, because even a more limited claim that without sentiments one cannot deliberate *well* is still in opposition to rationalism. This is because rationalists are best understood as not just claiming that reasoning alone is all we need for deliberating, but also as claiming that reasoning alone is all we need for deliberating well. Rationalists do not want to say that we can do some so-so deliberating through reasoning alone, but we'll do better if we involved sentiments; in fact, they want to say the opposite. So, I could concede the point about enhancing deliberation, and retain a distinctive non-rationalistic theory.

There is, however, a way of disarming the objection. I can deny that Spock can deliberate practically even though he possesses evaluative concepts with correct extensions. In order for this response to work, I need to say what practical deliberation consists in. I think it will be agreed on all sides that just thinking about actions does not count: thinking about the actions of a fictional character is not practical deliberation. Deliberation is practical if it is directed at answering the question of what I should do, i.e. practical deliberation must be capable of leading to rationally formed intentions. This statement of a necessary condition for practical deliberation is intuitive, and, I think, fairly uncontroversial. I shall now show that Spock's deliberation before he gains mastery fails to satisfy this condition. I have compared Spock's situation to that of a psychopath. Psychopaths may have moral concepts, yet with normative force bracketed: a psychopath may have the concept of what a moral thing to do is, but she is not motivated by this. Spock, similarly, has evaluative concepts, yet without motivational pull. Spock knows that other people are motivated by values, yet he himself is not motivated, in the same way that Mary knows what red is (in that she possesses the concept with the correct extension), yet she has not seen it herself. So,

[108] I owe this objection to Prof. Pink.

even though Spock can correctly apply the term 'good' to a given course of action, he is not motivated by it because the concept is so applied. If being capable of rational motivation is a necessary condition for practical deliberation, then Spock cannot deliberate practically: he cannot rationally form an intention to follow a particular course of action. Knowing perfectly well what other people call 'good' and being able to apply the term to the things that it applies to does not automatically bring motivation.[109] This response relies on a claim that I have not argued for explicitly, i.e. the claim that one has to have sentiments in order to be motivated. I am not going to provide a thorough-going defence of this claim here, because, as noted above, the objection is not a serious one, it just shows that my claims are unusual for a sentimentalist, and, secondly, because I am not entirely unhappy to retreat to the empirical level. Maybe Spock can deliberate in the absence of mastery, but we, human beings, cannot. As the case of psychopaths shows, people who have no mastery of evaluative concepts are not motivated by them. Moreover, in the human case, sentiments are necessary even for gaining a concept with the correct extension. This is illustrated by the fact that psychopaths, who lack what one may call 'moral emotions' over-extend moral concepts to conventional ones. CIP sufferers, mentioned at the beginning of this section, also help to illustrate the same point: they find it very difficult to keep track of dangerous things, and they also have some difficulty in attributing pain to others. So, if the reader is not convinced about what Spock can and cannot do (he is, after all, a being who is never inconsistent, never forgets anything, etc.), she should still accept that creatures like us are unable to deliberate practically without mastery of the relevant evaluative concepts.[110]

6.3. Mud, dirt, hair – the scope of the theory[111]

One question about any theory is its scope: which things does it apply to? This is one of the questions taken up in Plato's *Parmenides*. In this dialogue, Parmenides criticizes Plato's theory of forms. The theory is, very roughly, that there are abstract objects – the forms – that explain the values that we encounter. The form of Beauty explains why

[109] There may be some people who would say that if Spock is not motivated, then he does not have the concept 'good' at all. For reasons discussed in section 5.2.2.2., I think this is too strong. I do, however, agree with a weaker version of this claim: if Spock is *never* motivated by the good, then he does not have *mastery* of the concept 'good'.

[110] I say 'relevant' evaluative concepts because psychopaths can engage in practical deliberation in general, they just seem unable to deliberate practically about moral matters (as long as we accept that ability to form moral intentions is a necessary condition for moral practical deliberation).

[111] I thank my examiners, Dr Lillehammer and Prof. Wolff, for the comments that lead to a fuller discussion of this point and of the overall structure of the thesis.

beautiful boys, wise men, beautiful pieces of knowledge are all beautiful. Parmenides then asks Socrates:

> 'Is there a form, itself by itself, of just, and beautiful, and good, and everything of that sort?'
>
> 'Yes' he said.
>
> …
>
> 'And what about these, Socrates? Things that may seem absurd, like hair and mud and dirt, or anything else totally undignified and worthless? Are you doubtful about whether or not you should say that a form is separate for each of these, too, which in turn is other than anything we touch with our hands?' (*Parm*. 130b-d.)

What Parmenides wants to know is the scope of Socrates' proposal. Socrates says that his theory does not apply to such things, because they don't require explanation the way values do.

My critic, in a similar spirit, could question the scope of my theory. It may be true, they could say, that in order to have conceptual mastery of, say, 'rude' or 'unjust', one has to have the relevant experience. One cannot, to adapt Foot's (1958) example, fully grasp what is rude without ever having been offended, nor can one fully understand what is unjust without ever having felt indignation. But the need for experience might disappear when we get to a high enough level of abstraction – to such thin concepts as 'good' or 'moral'. The same goes for 'colour' in the Mary case. Oxford Dictionary Online defines colour as 'the property possessed by an object of producing different sensations on the eye as a result of the way it reflects or emits light'. It is not clear why Mary before her release cannot understand this definition as well as anyone who has seen colours. I have talked about phenomenal concepts as requiring to have had the relevant experience. It may be plausible for more specific concepts, such as 'indignant', but what is the relevant experience in the case of general concepts, such as 'good'? My response is that understanding of genera is enhanced by understanding of species. This claim can be cashed out by reference to the identification and application of generic concepts, but, as I show below, one cannot use the issue of application in order to answer the objection whilst staying neutral on metaphysics of value, which is what I would prefer to do.

Identification

The first issue is that Mary and Spock will be unable to identify even general, rather than particular, colours and values without the help of test subjects. It is true that Mary before her release can understand the definition of colour given above. Yet it lacks depth, as the following demonstrates. Mary, let us suppose, has conceptual mastery of 'colour', but not of individual colours. Spock, let us suppose, has conceptual mastery of 'good', but not of 'injustice', 'deliciousness', etc. Faced with an unfamiliar object, Mary would not be able to identify whether it is coloured without her instruments.[112] Her instruments must be calibrated with the help of test subjects who have mastery of colour concepts. Faced with an unfamiliar action, Spock would not be able to identify it as good without some test subjects, whose reactions he can observe – after all, the action's features fail to evoke a typical response in him. To enhance the parallel between that and the Mary case, we can give Spock an instrument that identifies which sentimental response a subject who possess sentiments would make. But this method is just a more indirect way of getting test subjects, since such subjects would be needed for the calibration of Spock's instruments.[113] Given this issue with identification, it is not clear that Mary and Spock can have conceptual mastery even of generic concepts, species of which require undergoing certain experiences. (Mary here may be in a better position than Spock – she can, after all, make colour judgements that are at least extensionally correct. It is less clear that Spock's judgements will be extensionally correct. As the study of psychopaths discussed in the previous chapter suggests, even mastering such general concepts as 'moral' requires a wide range of sentiments, at least in humans. Spock, however, is not human, so he may possess value concepts with correct extensions.)

Application

The second issue is that of application. Suppose we admit that Spock (much the same would apply to Mary) can make extensionally correct value judgements. We can also concede to the opponent that Spock has mastered general concepts such as 'good' or 'moral' – after all, these concepts are less clearly connected to sentiments. Would Spock

[112] To give plausibility to this situation, Mary must be colour blind, rather than confined to a black and white room. Otherwise she will gain mastery of the colour concept immediately on seeing the coloured object.

[113] One has to be careful in picking the right test subjects for calibrating one's instruments. Colour blind people will be unsuitable for Mary, psychopaths – for Spock. It is not clear to me that Mary and Spock can complete even the task of picking suitable participants without the help of someone who already has mastery of colour or evaluative concepts, respectively.

be able to apply the generic concepts when faced with particular cases? Can one grasp that a brave act is good without mastery of the concept 'brave'? A negative answer to this question requires a metaphysical commitment about values. Hurley (1989) is one making such a commitment to non-centralism, i.e. the idea that the general concepts (such as 'colour' and 'right') and specific ones (such as 'red' and 'just') are interdependent. Neither is prior to the other, but when they are grasped, they must be grasped together. Here is a short summary of Hurley's complex argument, which makes parallels between ethics, philosophy of mind, and philosophy of language. Meaning, evaluation and intentional action are not self-interpreting; these activities are only possible within some interpretation. But any interpretation must involve both formal and substantive constraints. Hurley argues – successfully, in my view – against a pervasive tendency to smuggle in substantive assumptions under the guise of formal requirements. So, when we are evaluating something (i.e. engaging in a particular type of interpretative activity), we have substantive ethical constraints which are interdependent with our specific evaluation. They are interdependent because without one or the other we would not be evaluating. This is essentially connected with Wittgensteinian idea of an 'austere conception of objectivity' (Hurley 1989, p. 226), i.e. the secondary-quality metaphysics of value, also popularized by McDowell (1998). This explanation is not available to me if I am to stay metaphysically neutral. Still, the previous point – that Spock cannot identify even generic values without the help of test subjects – holds, and requires no metaphysical commitments to what values are.

There is another way of pressing what is essentially the same objection. I have argued that evaluative concepts are phenomenal concepts: in order to master an evaluative concept, I must have had the relevant experience. But what is 'the relevant experience'? This is a fair question, yet I cannot answer it here to my satisfaction. This is because we lack a taxonomy both of sentiments and of evaluative concepts. Contemporary philosophical discussion of the latter tends to concentrate on generic evaluative concepts, such as 'right', 'ought', and 'moral'. Not many authors (Hurley is an exception) explicitly discuss the relationship between different evaluative concepts.[114] Something similar can be said about the taxonomy of sentiments. A lot of work has been done on

[114] Discussions about the unity of virtue – i.e. Aristotle's thesis that someone who possesses one virtue will necessarily possess them all (*NE* 1144b-1145a2) – may be a helpful place to start when trying to provide a taxonomy of moral concepts. According to this idea, if not all evaluative concepts, then at least moral ones are interrelated in such a way that you cannot master them separately. Once you fully grasp one moral concept, you fully grasp its relationships with other moral concepts, too. However, as Badhwar (1996) points out, the idea of the unity of virtue is not uncontroversial: Williams, for example, takes its rejection, i.e. the claim that one can have one virtue without having all of them, a platitude (1985, p. 36).

distinguishing 'the basic emotions'. Ekman (1989) lists happiness, sadness, fear, anger, surprise, and disgust, but this list is not universally accepted. Moreover, I have spoken not about emotions, but about a broader category of sentiments, so a classification of emotions is going to be only one part of the task. We need a taxonomy of evaluative concepts, a taxonomy of sentiments, and only then it is possible to work out the relationship between the two in satisfying detail. This, obviously, is a huge task, but, in my view, a necessary one for *any* sentimentalist account to be credible. Empirical studies will be of help here – e.g. autistic children lack empathy, yet are moral (e.g. Blair (1996) showed that autistic children do make the moral/conventional distinction). I have provided at least one case study in support of my view: as I argued in Chapter 3, psychopaths do fail to master moral concepts, which is best explained by a deficit in emotions such as guilt and remorse.

There is also new empirical evidence that, in humans, even such abstract concepts as permissibility require sentiments.[115] Patients with damage to the ventromedial pre-frontal cortex (VMPFC), discussed in the previous chapter, also appear to have an unusual (by comparison to those whose emotions are normal) extension of a particular type of evaluative concepts: moral concepts.[116] Koenigs *et al.* (2007) found that patients with VMPFC damage have abnormal moral judgements: unlike controls, they judge personal moral harms to be permissible. This is illustrated by the trolley/footbridge cases. In the trolley case, you can save five lives by flipping a switch, so that a runaway trolley changes direction; this would, however, result in the death of one person on another track. The footbridge scenario is the same, apart from the fact that you stop the trolley from reaching the five people by pushing a man off the footbridge. Once he is

[115] This complements the discussion in the previous chapter. There, I concentrated on two particular studies – Iowa Gambling Task and Blair's study of the moral/conventional distinction, because both have been widely discussed in the literature. However, I argued that the Iowa Gambling Task does not actually show what Damasio and other sentimentalists take it to show, i.e. that the best explanation of patients' behaviour is that they lack somatic markers. The aim of discussion there was to invite a certain caution first, about interpretation of studies, and, secondly, about presenting the Iowa Gambling Task as uncontroversially supporting sentimentalism; as we have seen, an equally good rationalist explanation of the study's results is available. Blair's study of the moral/conventional distinction, whilst unquestionably philosophically interesting, has recently become controversial. Dolan and Fullam (2010) have only partially replicated Blair's results. They found that the only dimension on which psychopaths in young offenders' institutions failed to distinguish between moral and conventional transgressions was authority-dependence. On other dimensions, such as seriousness, permissibility and justification, the performance of those with high psychopathic tendencies did not differ significantly from controls. Moreover, Aharoni *et al.* (2011) found that prisoners with high psychopathy scores were no worse than controls at distinguishing between moral and conventional norm violations. (Although Aharoni *et al.*'s study used a different method from Blair's.) Thus, this discussion brings in more recent evidence of psychopaths' moral abilities.

[116] I am interested in evaluative concepts in general, rather than moral concepts in particular. However, I concentrate on moral concepts here because they are well-researched and because it is less obvious that sentiments are required for mastery for such abstract concepts (cf. 'tasty').

pushed onto the tracks, his body stops the trolley. Normally, people say it is permissible to stop the trolley by using the switch, but impermissible to stop it by pushing the man onto the tracks. People with VMPFC damage tend to say that both are permissible. Researchers hypothesize that this is because VMPFC patients' emotions are impaired: contemplating pushing a person off the bridge 'would normally evoke a strong social emotion', leading to judgement of impermissibility' (Koenigs *et al.* 2007, p. 910). VMPFC patients have dulled emotions (participants in this particular study had impaired empathy, embarrassment and guilt), so they fail to think that the action is impermissible:

Normal pattern: personal harm – emotional response – it is not permissible to do this.

VMPFC patients' pattern: personal harm – no emotional response – it is permissible to do this.[117]

Other studies support the idea that these judgements are due to emotional deficits. Ciaramelli *et al.* (2007) replicated this result, as well as showing that it was due to emotional rather than cognitive deficits (such as impulsiveness or working memory problems). Martins *et al.* (2012) studied patients with traumatic brain injury, whose lesions were not constrained to VMPFC area. They found a positive correlation between difficulties in emotional processing and rating personal harms as more permissible.

Thus, patients with VMPFC damage over-extend the concept 'permissible actions', and they do so because their emotions are impaired. These agents can still make moral judgements – they provide normal responses to the trolley problem – but they have a selective deficit, as shown by deviant responses about personal harms. And this shows that emotional experiences play a role in determining extension of at least some abstract moral concepts. Further evidence for the same claim comes from studies of psychopaths' moral judgements.

Psychopaths' moral judgements diverge from those made by non-psychopaths. First, they judge accidental harms as more morally permissible. For example, if I put some white grainy substance that is labelled 'sugar' into a colleague's coffee cup, and my colleague dies as a result (because, as it turns out, the substance was highly toxic),

[117] I do not mean that VMPFCs patients' judgements are incorrect. All that is important for my purposes is that their moral judgement differs from normal because of deficiencies in emotional experience. This shows that our moral concepts depend on ability to experience emotions. This admission opens up an option that is commonly neglected: that patients with VMPFC damage are moral, but their morality is different from ours. The studies' results are them taken to show that a moral concepts of normal, non-brain-damaged humans, are dependent on normal sentiments.

psychopaths tend to say that this is permissible, whilst non-psychopaths tend to say it is not (Young *et al.* 2012). Secondly, psychopaths are more likely than non-psychopaths to endorse impersonal harms.[118] For example, most people say it is ok to divert the trolley by pushing the switch, yet psychopaths do so more readily than others. Psychopaths with low anxiety also make the same judgements as patients with VMPFC damage in personal harm cases, such as the footbridge case.

These studies show that even in the case of abstract moral concepts – the traditional domain of rationalists – sentiments are important, since human agents with abnormal sentiments have deviant extensions of the concept of permissibility.

So far I have presented a thesis that is fairly general: in order to master evaluative concepts, one has to have had certain sentiments. However, there are different ways of making this thesis more substantial. This may be necessary in order to clarify the relationship between the arguments presented in this chapter and the previous one. Suppose, for example, that values are natural properties of things. In this case, it is not clear why Spock cannot make masterful evaluations even though he lacks sentiments, contrary to the conclusion of the argument in section 5.2. There are several things I can do in order to answer this objection.

First, instead of staying neutral on the metaphysics of value, I could, like McDowell (1985) and Prinz (2007) liken values to secondary qualities.[119] If a virtuous action cannot be understood except in relation to the sensibilities of those who classify it as such, Spock, who does not share this sensibility, will be unable to master the concept of virtue. Virtuous actions will be as much of a mystery to him as the distinctive phenomenology of colours is to the colour blind. If this option is taken, the arguments presented in sections 5.2 and 5.1 of this chapter are the main thrust of the thesis about mastery, whilst real cases described in Chapter 3 play supporting, illustrative role.

I am not attracted to this option because it entails what Lillehammer calls 'an air of relativism' (Lillehammer 2011, p.64). That is, the secondary-quality view of values allows for as many competing evaluative outlooks as there are differing sensibilities,

[118] These results directly contradict those of Cima *et al.* (2010), who found no differences between psychopaths' and non-psychopaths' patterns of judgement in these cases. Koenigs *et al.* (2012) explain this discrepancy by the difference in what they and Cima *et al.* defined as 'psychopathy': Koenigs *et al.* used a higher cutoff score for psychopathy. Indeed, when Koenigs *et al.* used the same score as Cima *et al.*, they found no difference in judgement, either. This suggests that someone with psychopathic tendencies, but not classed as a psychopath, may exhibit the normal pattern of moral judgement, whereas someone classed as a psychopath does not.

[119] In fact, this metaphysical commitment may be considered necessary for my thesis to go through. I rebut this suggestion in note 103 above.

each of these outlooks being as justified as another.

Secondly, one could turn to language, and argue that sentiments are part of the meaning of evaluative concepts (e.g. Blackburn 1998, Prinz 2007). According to this view, one does not know what 'right' is unless one (always, or at least typically) has some negative feeling when one uses this concept in a sincere assertion in a certain type of proposition.[120] If this option is taken, the arguments of this thesis play out in the same way as in the case of secondary-quality theory of value, with the arguments about empirical studies playing supplementary role.

There are several problems associated with this option, two of them already discussed. First, I can make dispassionate evaluations (section 5.2.2.2.), although this problem is avoided by the weaker version of the view, which requires that I have the sentiment only *typically*. Secondly, the feelings one has when using concepts are closely connected with motivation. Thus, the Too Many and Too Few Reasons problems, discussed in Chapter 2, re-appear. Thirdly, this option is usually coupled with an even more austere metaphysics than the secondary-quality theory of value: one's values are taken to be constituted by the sentiments one has. This allows for even more competing theories of value, criticisms of which are a matter of personal preference.

Instead of turning to metaphysics or to a thesis about meaning of evaluative terms, one could restrict the mastery thesis to a certain type of agents. The mastery thesis may fail to apply to perfectly rational aliens like Spock, but it applies to agents who are human and sufficiently similar.[121] I am attracted to this last option, since it avoids the problems of the first two. It does not require a link between motivation and evaluation that I find objectionably strong. Nor does it exclude the possibility of values as primary qualities. This possibility is required for robust justification and criticism of competing evaluative outlooks.[122] If this option is taken, then empirical arguments of the previous chapter provide the main support for the mastery thesis, whilst the arguments of section 5.2, which requires a commitment to a specific metaphysics of value, does not go through.

---

[120] ' A certain type' means that questions, conditionals, etc. are excluded, and what one may call 'basic' moral propositions, such as 'Murder is wrong.', are included.

[121] We find out whether an agent is sufficiently similar to us by running the experiments to establish whether these agents' evaluative concepts change extension and whether they lead to motivation if the agents' sentiments are abnormal.

[122] I do assume that we need to have objective values in order to justify our pre-philosophical talk about good reasons. But I had little chance to explore the irrealist cognitivist suggestion (Skorupski 1999) that one should, instead, settle on the question of what truth is first. Accepting a minimalist theory of truth might allow us to justify fully our talk of good reasons without there being any.

Obviously, empirical arguments are only as good as the current empirical evidence, but that is an acceptable, if not a necessary, risk when one is primarily interested in human agents, rather than agents *per se*.

7. Conclusion

I have argued that someone who lacks sentiments will be unable to gain mastery of evaluative concepts by making a parallel with the Knowledge Argument in philosophy of mind. I have also shown why this lack of mastery is important: such a being will be unable to have (access to) normative reasons for action.

Conclusion

My question has been: what sort of beings respond to good reasons for action? In so far as we are such beings, the question is about us, humans. I have argued that, in order for us to respond to good reasons, we need sentiments. This is so because sentiments provide mastery of evaluative concepts, which then figure in our evaluations. Depending on which metaphysics one has, these evaluations either constitute one's normative reasons for action or represent them.

This answer belongs to a sentimentalist camp. I argued for sentimentalism by showing, in Chapter 1, that a purely formal rationalist theory fails to explain how we respond to good reasons. There I have argued that a rationalist theory must say that one must be equally motivated to follow equally consistent but opposing courses of action. Sentimentalism is also supported by empirical studies – the existence of psychopaths, who are rational yet amoral, is best explained by a sentimentalist theory (Chapter 3). I have also shown that my brand of sentimentalism overcomes traditional problems (Chapters 2 and 4). In doing so, I have provided a distinctive theory that links normative reasons for action with sentiments. This link, as I have shown, is best understood not in terms of having reasons to do what promotes my desires, but in terms of mastery of evaluative concepts.

With this I answer the question posed at the start: what sort of beings do we have to be in order to do things, and often for good reasons? Such beings must have a capacity for sentiment; a capacity for reasoning alone will not do the job.

Bibliography

Aharoni, E., Sinnott-Armstrong, W., and Kiehl, K. (2011). 'Can Psychopathic Offenders Discern Moral Wrongs? A New Look at the Moral/Conventional Distinction', *Journal of Abnormal Psychology*, 121:2, 484-497.

Alter, T. (2008). 'Phenomenal Knowledge without Experience' in Wright, E. (ed.) *The Case for Qualia*. MIT Press, 247-267.

Alter, T. (2011). 'Tye's New Take on the Puzzles of Consciousness', *Analysis Reviews*, 71:4, 765–775.

Alter, T. (forthcoming). 'Social Externalism and the Knowledge Argument', *Mind*.

Alter, T., and Walter, S. (eds.) (2007). *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford: Oxford University Press.

Anscombe, G.E.M. (1957, edition referred to 2000). *Intention*. Harvard University Press: Cambridge, Massachusetts.

Aristotle, *Nicomachean Ethics*, Ross, W.D. (trans.) Available: http://classics.mit.edu/Aristotle/nicomachaen.html. Last accessed 16 February 2012.

Ayer, A. J. (1936, edition referred to 2002). *Language, Truth, and Logic*. Dover Publications Inc.

Babiak, P. (2007). 'From Darkness Into the Light: Psychopathy in Industrial and Organisational Psychology' in Herve, H. and Yuille, J.C. (eds.) (2007), 411–428.

Badhwar, N. K. (1996). 'The Limited Unity of Virtue', *Noûs*, 30:3, 306-329.

Ball, D. (2009). 'There Are No Phenomenal Concepts', *Mind*, 118, 935-962.

Baron-Cohen, S., Leslie, A., and Frith, U. (1985). 'Does the Autistic Child Have a "Theory Of Mind"?' *Cognition*, 21, 37-46.

Beaton, M. (2005). 'What RoboDennett Still Doesn't Know', *Journal of Consciousness Studies*, 12, 3-25.

Bechara, A., Damasio, A. R., Damasio, H., and Anderson, S.W. (1994). 'Insensitivity to Future Consequences Following Damage to Human Prefrontal Cortex', *Cognition*, 50, 7-15.

Bechara, A., Damasio, H., Tranel, D., and Damasio, A.R. (1997). 'Deciding Advantageously before Knowing the Advantageous Strategy', *Science*, 275, 1293-1295.

Bechara, A., Damasio, H., Tranel, D., and Damasio, A.R. (2005). 'The Iowa Gambling Task and the Somatic Marker Hypothesis: Some Questions and Answers', *Trends in Cognitive Sciences*, 9:4, 159-162.

Biggs, S. (2009). 'Phenomenal Concepts in Mindreading', *Philosophical Psychology*, 22:6, 647-667.

Blackburn, S. (1984). *Spreading the Word*. New York: Oxford University Press.

Blackburn, S. (1998). *Ruling Passions.* Clarendon Press: Oxford.

Blair, J. (1995). 'A Cognitive Developmental Approach to Morality: Investigating the Psychopath', *Cognition*, 57, 1-29.

Blair, J. (1996). 'Brief Report: Morality in the Autistic Child', *Journal of Autism and Developmental Disorders*, 26:5, 571-579.

Blair, J. (1997). 'Moral Reasoning and the Child with Psychopathic Tendencies', *Personality and Individual Differences,* 22:5, 731-739.

Blair, J. (2008). 'The Amygdala and Ventromedial Prefrontal Cortex: Functional Contributions and Dysfunction in Psychopathy', *Philosophical Transactions of the Royal Society*, 363, 2557–2565.

Blair, J., Jones, L., Clark, F., and Smith, M. (1995). 'Is the Psychopath "Morally Insane"?', *Personality and Individual Differences*, 19:5, 741-752.

Blair, J., Mitchell, D., and Blair, K. (2005). *The Psychopath: Emotion and the Brain.* Blackwell Publishing.

Blair, J., Monson, J., and Frederickson, N. (2001). 'Moral Reasoning and Conduct Problems in Children with Emotional and Behavioural Difficulties', *Personality and Individual Differences*, 31, 799-811.

Broome, J. (1999). 'Normative Requirements', *Ratio*, 12, 398–419.

Budhani, S., Richell, R.A., and Blair, R. J. R. (2006). 'Impaired Reversal but Intact Acquisition: Probabilistic Response Reversal Deficits in Adult Individuals with Psychopathy', *Journal of Abnormal Psychology*, 115:3, 552-558.

Burge. T. (1979). 'Individualism and the Mental', *Midwest Studies in Philosophy*, 4:1, 73-121.

Calder, A.J., Keane, J., Lawrence A. D., and Manes, F. (2004). 'Impaired Recognition of Anger Following Damage to the Ventral Striatum', *Brain*, 127, 1958–1969.

Ciaramelli, E., Muccioli, M., La davas, E., and di Pellegrino, G. (2007). 'Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex', *Social, Cognitive, and Affective Neuroscience*, 2, 84–92.

Cima, M, Tonnaer F., and Hauser, M. D. (2010). 'Psychopaths know right from wrong but don't care', *Social, Cognitive, and Affective Neuroscience*, 5, 59-67.

Cleckley, H. (1988). *The Mask of Sanity.* Fifth Edition: private printing for non-profit educational use, available: www.cassiopaea.org/cass/sanity_1.PdF.

Coleman, S. (forthcoming). 'Review of *Consciousness Revisited* by Michael Tye', *Philosophy.*

Colombetti, G. (2008). 'The Somatic Marker Hypotheses, and What the Iowa Gambling Task Does and Does not Show', *British Journal of Philosophy of Science*, 59, 51-71.

Crane, T. (2005). 'Papineau on Phenomenal Concepts', *Philosophy and Phenomenological Research*, 71:1, 155-162.

Damasio, A.R. (1994). *Descartes' Error: Emotion, Reason and the Human Brain.* Avon: New York.

Damasio, A.R. (1996). 'The Somatic Marker Hypothesis and the Possible Functions of the Prefrontal Cortex', *Philosophical Transactions: Biological Sciences*, 351, 1413-1420.

Damasio, A.R., Grabowski, T.J., Bechara, A., Damasio, H., Ponto, L.L., Parvizi, J., and Hichwa, R.D. (2000). 'Subcortical and cortical brain activity during the feeling of self-generated emotions', *Nature Neurosience*, 3, 1049-1056.

Damasio, A.R., Tranel, D., and Damasio H. C. (1991). 'Somatic Markers and the Guidance of Behaviour' in Jenkins, J.M., Oatley, K. and Stein, N.L. (eds.) (1998) *Human Emotions:A Reader*. Blackwell publishers, 122-135.

Dancy, J. (2000). *Practical Reality*. Oxford: Oxford University Press.

Danzinger, N., Prkachin, K., and Willer, J. (2006). 'Is Pain the Price of Empathy? The perception of others' pains in patients with congenital insensitivity to pain', *Brain*, 129, 2594–2507.

Dennett, D. (1993). 'The Message is: There is no Medium (reply to Jackson, Rosenthal,

Shoemaker and Tye)', *Philosophy and Phenomenological Research*, 53:4, 889-931.

Dennett, D. (1995). 'Do Animals Have Beliefs?' in Roitblat, H. (ed.) *Comparative Approaches to Cognitive Sciences*. MIT Press.

Dennett, D. (2007). 'What RoboMary Knows,' in Alter, T. and Walter, S. (eds.) (2007), 15–31.

Dolan, M.C. and Fullam, R. S. (2010). 'Moral/conventional transgression distinction and psychopathy in conduct disordered adolescent offenders', *Personality and Individual Differences*, 49, 995-1000.

Dunn, B. D., Dalgleish, T., and Lawrence, A.D. (2006). 'The Somatic Marker Hypothesis: A Critical Evaluation', *Neuroscience and Biobehavioral Reviews*', 30, 239–271.

Ekman, P. (1989). 'The Argument and Evidence About Universals in Facial Expressions of Emotion,' in Wagner, H. and Manstead, A. (eds.) *Handbook of Social Psychophysiology*. Chichester, England: Wiley, 143-164.

Fellows, L.K. and Farah, M.J. (2005). 'Different Underlying Impairments in Decision-Making Following Ventromedial and Dorsolateral Frontal Lobe Damage in Humans', *Cerebral Cortex*, 15:1, 58–63.

Figner, B. and Murphy, R. O. (2011). 'Using Skin Conductance in Judgment and Decision Making Research', in Schulte-Mecklenbeck, M., Kuehberger, A. and Ranyard, R. (eds.) *A Handbook of Process Tracing Methods for Decision Research*. New York, NY: Psychology Press, 163-184.

Foot, P. (1958). 'Moral Arguments' in her *Virtues and Vices* (1978). Basil Blackwell, 96 -109.

Foot, P. (1972). 'Morality as a System of Hypothetical Imperatives', *The Philosophical Review*, 81:3, 305-316.

Frankfurt, H. G. (1971). 'Freedom of the Will and the Concept of a Person', *The Journal of Philosophy*, 68:1, 5-20.

Gao, Y. and Raine, A. (2010). 'Successful and Unsuccessful Psychopaths: A Neurobiological Model', *Behavioural Sciences and the Law*, 28, 194-210.

Geach, P. (1965). 'Assertion', *The Philosophical Review*, 74:4, 449-465.

Glenn, A.L, Iyer, R., Graham, J., Koleva, S., and Haidt, J. (2009). 'Are All Types of Morality Compromised in Psychopathy?', *Journal of Personality Disorders,* 23:4, 384-

398.

Gregory, A. (2009). 'Slaves of the Passions? On Schroeder's New Humeanism', *Ratio (new series)*, 22, 250-257.

Goldie, P. (2000). *The Emotions*. Oxford: Clarendon Press.

Goldie, P. (2006). 'Review: "Gut Reactions: A Perceptual Theory of Emotion,"' *Mind*, 115:458, 453-457.

Haidt, J., Koller, H. S., and Dias, M. G. (1993). 'Affect, Culture and Morality, or Is It Wrong to Eat Your Dog?', *Journal of Personality and Social Psychology*, 65:4, 613-628.

Hale, B. (2002). 'Can Arboreal Knotwork Help Blackburn out of Frege's Abyss?', *Philosophy and Phenomenological Research*, 65:1, 144-149.

Hampshire, S. (1999). ' "The Reason Why Not", Review of Scanlon 1999', *New York Review of Books*, 22 Apr 1999.

Hare, R.D. (1991). *The Hare Psychopathy Checklist–Revised*. Toronto, Ontario: Multi-Health Systems.

Hare, R.D. (1993). *Without Conscience: the Disturbing World of Psychopaths among Us*. The Guilford Press: New York, London.

Hare, R.D. (2007). 'Forty years Aren't Enough: Recollections, Prognostications, and Random Musings' in Herve, H. and Yuille, J.C. (eds.) (2007), 3-30.

Herve, H. and Yuille, J.C. (eds.) (2007). *The Psychopath Theory: Research and Practice*. Lawrence Erlbaum Associates Publishers.

Heims, H.C., Critchley, H.D., Dolan, R., Mathias, C.J., and Cipolotti, L. (2004). 'Social and Motivational Functioning is Not Critically Dependent on Feedback of Autonomic Responses: Neuropsychological Evidence from Patients with Pure Autonomic Failure', *Neuropsychologia*, 42, 1979–1988.

Howell, R. J. (2011). 'The Knowledge Argument and the Implications of Phenomenal Knowledge' , *Philosophy Compass*, 6, 459–468.

Hume, D. (1738-1740, edition referred to 1978). *A Treatise of Human Nature*. Selby-Bigge, L. A. (ed.) 2nd edition revised by Nidditch, P.H. (1978). Oxford: Clarendon Press.

Hume, D. (1777, edition referred to 1975). *Enquiry concerning Human Understanding*

in *Enquiries concerning Human Understanding and concerning the Principles of Morals*, Selby-Bigge, L.A., 3rd edition revised by Nidditch, P.H. (1975). Oxford: Clarendon Press.

Hurley, S. (1989). *Natural Reasons*. Oxford University Press: New York, Oxford.

Ishikawa, S., Raine, A., Lencz, T., Bihrle, S., and Lacasse, L. (2001). 'Autonomic Stress Reactivity and Executive Functions in Successful and Unsuccessful Criminal Psychopaths From the Community', *Journal of Abnormal Psychology*, 110:3, 423-432.

Izquierdo, A., Suda, R. K., and Murray, E. A. (2005). 'Comparison of the Effects of Bilateral Orbital Prefrontal Cortex Lesions and Amygdala Lesions on Emotional Responses in Rhesus Monkeys', *The Journal of Neuroscience*, 25:37, 8534 – 8542.

Izquierdo, A. and Murray, E. A. (2007). 'Selective Bilateral Amygdala Lesions in Rhesus Monkeys Fail to Disrupt Object Reversal Learning', *The Journal of Neuroscience*, 27:5, 1054-1062.

Jackson, F. (1982). 'Epiphenomenal Qualia', *Philosophical Quarterly*, 32, 127–136.

Jackson, F. (1998). 'Postscript on Qualia' in his *Mind, Methods and Conditionals* (1998). London: Routledge, 76-79.

Jones, A. P., Happé, F. G. E., Gilbert, F., Burnett, S., and Viding, E. (2010). 'Feeling, Caring, Knowing: Different Types of Empathy Deficit in Boys with Psychopathic Tendencies and Autism Spectrum Disorder', *Journal of Child Psychology and Psychiatry*, 51:11, 1188-1197.

Kant, I. (1785, edition referred to 1964). *Groundwork for the Metaphysics of Morals*, Paton, H. J. (trans.) (1964). Harper Perennial.

Kant, I. (1788, edition referred to 1976). *Critique of Practical Reason* in Beck, L.W. (ed. and trans.) (1976) *Critique of Practical Reason and Other Writings in Moral Philosophy*. The University of Chicago Press: Chicago, Illinois.

Kant, I. (1797, edition referred to 1976). 'On a Supposed Right to Lie from Altruistic Motives' in Beck, L.W. (ed. and trans.) (1976) *Critique of Practical Reason and Other Writings in Moral Philosophy*. The University of Chicago Press: Chicago, Illinois.

Kennett, J. (2006). 'Do psychopaths really threaten moral rationalism?', *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, 9:1, 69-82.

Koenigs, M, Kruepke, M., Zeier, J., and Newman, J. P. (2012). 'Utilitarian moral

judgment in psychopathy', *Social, Cognitive, and Affective Neuroscience*, 7, 708-714.
Koenigs, M, Young, L, Adolphs, R, Tranel, D., Cushman, F, Hauser, M., and Damasio, A. (2007). 'Damage to the prefrontal cortex increases utilitarian moral judgements', *Nature*, 446:7138, 908-911.

Korsgaard, C. (1986, edition referred to 1998). 'The Right to Lie: Kant on Dealing with Evil' in Rachels, J. (ed.) *Ethical Theory 2: Theories About How We Should Live* (1998). Oxford: Oxford University Press, 282-304.

Korsgaard, C. (1996). *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.

Korsgaard, C. (1997, edition referred to 2008). 'The Normativity of Instrumental Reason' in her *The Constitution of Agency, Essays on Practical Reason and Moral Psychology* (2008). Oxford: Oxford University Press, 27-68.

Lambert, K. (2007). 'How CIPA Works', available: HowStuffWorks.com. <http://science.howstuffworks.com/environmental/life/human-biology/cipa.htm>, accessed 09 August 2012.

Lau, J. and Deutsch, M. (2010). 'Externalism About Mental Content', *The Stanford Encyclopedia of Philosophy (Fall 2010 Edition)*, Zalta, E. N. (ed.), URL = <http://plato.stanford.edu/archives/fall2010/entries/content-externalism/>.

Levenson, R. W., Ekman, P., and Friesen, W. V. (1990). 'Voluntary Facial Action Generates Emotion-Specific Autonomic Nervous System Activity', *Psychophysiology*, 27, 363-384.

Levin, J. (2007). 'What Is a Phenomenal Concept?' in Alter, T. and Walter, S. (2007), 87-110.

Lewis, D. (1983). 'Postscript to "Mad Pain and Martian Pain"', in his *Philosophical Papers Volume 1* (1983). Oxford: Oxford University Press, 130-133.

Lewis, D. (1988). 'What Experience Teaches', *Proceedings of the Russellian Society*, 13, 29-57.

Lillehammer. H. (2007). *Companions in Guilt: Arguments for Ethical Objectivity*. London: Palgrave Macmillan.

Lillehammer, H. (2011). 'Constructivism and the error theory' in Miller, C. (ed.) *The Continuum Companion to Ethics* (2011). New York: Continuum, 55-76.

Loar, B. (1997). 'Phenomenal States' (Revised Version), in Block,N., Flanagan, O., and

Güzeldere, G. (eds.) *The Nature of Consciousness: Philosophical Debates*. Cambridge, MA: MIT Press, 597-616 (revised version of Loar 1990).

Locke, J. (1689, edition referred to 1997). *An Essay Concerning Human Understanding*. Woolhouse, R. (ed.) (1997). Penguin Books.

Mackie, J.L. (1977). *Ethics: Inventing Right and Wrong*. New York: Penguin.

Maia, T.V. and McClelland, J.L. (2004). 'A Reexamination of the Evidence for the Somatic Marker Hypothesis: What Participants Really Know in the Iowa Gambling Task', *Proceedings of the National Academy for Science of the United States of America*, 101:45, 16075-16080.

Maia, T.V. and McClelland, J.L. (2005). 'The Somatic Marker Hypothesis: Still Many Questions but no Answers: Response to Bechara *et al.*', *Trends In Cognitive Sciences*, 9:4, 162-164.

Maibom, H. (2005). 'Moral Unreason: A Case of Psychopathy', *Mind and Language*, 20:2, 237-257.

Martins, A. T., Faísca, L.M., Esteves, F., Muresan, A., Reis A. (2012). 'Atypical moral judgment following traumatic brain injury', *Judgment and Decision Making*, 7:4, July 2012, 478-487.

McDowell, J. (1978, edition referred to 1998). 'Are Moral Requirements Hypothetical Imperatives?', *Proceedings of the Aristotelian Society*, 52, 13-29.

McDowell, J. (1979,edition referred to 1998). 'Virtue and Reason', *The Monist*, 62, 331-350 reprinted in his (1998).

McDowell, J. (1985, edition referred to 1998). 'Values and Secondary Qualities' in Honderich, T. (ed.) *Objectivity and Morality* (1985). London: Routledge and Kegan Paul, 110-129, reprinted in his (1998).

McDowell, J. (1998). *Mind, Value and Reality*. Harvard University Press.

McIntyre, A. (1990). 'Is Akratic Action Always Irrational?,' in *Identity, Character, and Morality*, Flanagan, O. and Rorty, A. (eds.). Cambridge, MA: MIT Press, 379-400.

Morrisey, B. (updated: 26 April 2012). Available: http://www.eatingdisorderexpert.co.uk/picadisorder.html. Last accessed 19 July 2012.

Mullins-Nelson, J.L, Salekin, R.T., and Leistico, A. R. (2006). 'Psychopathy, Empathy, and Perspective-Taking Ability in a Community Sample: Implications for the

Successful Psychopathy Concept', *International Journal of Forensic Mental Health*, 5:2, 133-149.

Naccache, L., Dehaene, S., Cohen, L., Habert, M., Guichart-Gomez, E., Galanaud, D., and Willer, J-C. (2005). 'Effortless Control: Executive Attention and Conscious Feeling of Mental Effort Are Dissociable', *Neuropsychologia*, 43, 1318-1328.

Nagel, T. (1970). *The Possibility of Altruism*. Oxford: Clarendon Press.

Nemirow, L. (1980). 'Review of Thomas Nagel, *Mortal Questions*,' *Philosophical Review*, 89, 473-477.

Nemirow, L. (2007). 'So *This* Is What It's Like: A Defence of the Ability Hypothesis' in Alter. T. and Walter, S. (eds.) (2007), 32-51.

Newman, J.P., Brinkley, C.A., Lorenz A. R., Hiatt, K.D., and MacCoon, D.G. (2007). 'Psychopathy as Psychopathology: Beyond the Clinical Utility of the Psychopathy Checklist–Revised' in Herve, H. and Yuille, J.C. (eds.) (2007), 173-206.

Nida-Rumelin, M. (1996). 'What Mary Couldn't Know: Belief About Phenomenal States' in Metzinger, T. (ed.) *Conscious experience*. Exeter: Imprint Academic, 219–242.

Nichols, S. (2002a). 'Is it Irrational to Be Amoral? How Psychopaths Threaten Moral Rationalism', *The Monist*, 85, 285-304.

Nichols, S. (2002b). 'Norms with Feeling: Towards a Psychological Account of Moral Judgment', *Cognition*, 84, 221-236.

Nichols, S. (2004). *Sentimental Rules*. Oxford: Oxford University Press.

Noe, A. (2006). 'Experience Without the Head' in Gendler,T. S. and Hawthorne, J. (eds.) *Perceptual Experience*. Oxford: Oxford University Press, 411-433.

Oddie, G. (2005). *Value Reality and Desire*. Oxford: Clarendon Press.

O'Neill, O. (1985, edition referred to 1998). 'Consistency in Acton' in Rachels, J. (ed.) *Ethical Theory 2: Theories About How We Should Live* (1998). Oxford: Oxford University Press, 256-281.

Oxford Dictionaries (2012). Available: http://oxforddictionaries.com/. Last accessed 16 February 2012.

Papineau, D. (2002). *Thinking about Consciousness*. Oxford: Oxford University Press.

Papineau, D. (2007). 'Phenomenal and Perceptual Concepts' in Alter, T. and Walter, S.

(eds.) (2007), 111-144.

Patrick, C. J. (2007). 'Getting to the Heart of Psychopathy' in Herve, H. and Yuille, J.C. (eds.) (2007), 207-252.

Pink, T. (2004). 'Suarez, Hobbes and The Scholastic Tradition in Action Theory' in Pink,T. and Stone, M.W.F. (eds.) *The Will and Human Action* (2004). Routledge, 127-153.

Pink, T. (2008). 'Intentions and Two Models of Human Action' in Verbeek, B. (ed.) *Reasons and Intentions*. Ashgate, 153-181.

Pink, T. (2009a). 'Power and Moral Responsibility', *Philosophical Explorations*, 12:2, 127-149.

Pink, T. (2009b). 'Reason, Voluntariness, and Moral Responsibility' in O'Brien., L. and Soteriou, M. (eds.) *Mental Actions*. Oxford: Oxford University Press, 95-120.

Plato *Complete Works*, J.M. Cooper (ed.). Hackett Publishing Company: Indianapolis/Cambridge.

Porter, S. and Porter, S. (2007). 'Psychopathy and Violent Crime' in Herve, H. and Yuille, J.C. (eds.) (2007), 287-300.

Prinz, J. (2004). *Gut Reactions: A Perceptual Theory of Emotion*. Oxford: Oxford University Press.

Prinz, J. (2006). 'The Emotional Basis of Moral Judgements', *Philosophical Explorations*, 9:1, 29-43.

Prinz, J. (2007). *The Emotional Construction of Morals*. Oxford: Oxford University Press.

Quinn, W. (1993). 'Putting Rationality in its Place' in his *Morality and Action*. Cambridge Studies in Philosophy: Cambridge University Press, 228-255.

Rabin, G. (2011). 'Conceptual Mastery and the Knowledge Argument', *Philosophical Studies*, 154, 125-147.

Raz, J. (1986). *The Morality of Freedom*. Oxford: Clarendon Press.

Ridge, M (2006). 'Ecumenical Expressivism: Finessing Frege', *Ethics*, 116:2, 302-336.

Rolls, E.T. (1999). *The Brain and Emotion*. Oxford University Press.

Rolls, E.T. (2000). 'The Orbitofrontal Cortex and Reward', *Cerebral Cortex*, 10, 284-

294.

Rolls, E.T., Hornak, J., Wade, D., and McGrath, J. (1994). 'Emotion-related Learning in Patients with Social and Emotional Changes Associated with Frontal Lobe Damage', J*ournal of Neurology, Neurosurgery, and Psychiatry*, 57, 1518-1524.

Scanlon, T.M. (1998). *What We Owe to Each Other*. Belknap Press: Harvard University Press.

Schroeder, M. (2007). *Slaves of the Passions*. Oxford: Oxford University Press.

Schnall, S., Haidt, J., Clore,G. L., and Jordan, A.H. (2008). 'Disgust as Embodied Moral Judgment', *Personal and Social Psychology Bulletin*, 34:8, 1096-1109.

Schueler, G. F. (1995). *Desire: Its Role In Practical Reason And The Explanation Of Action*. MIT Press: Cambridge, Massachusetts.

Singer, M. (1961). *Generalization in Ethics*. New York: Russell and Russell.

Skorupski, J. (1999). 'Irrealist Cognitivism', *Ratio*, 12:4, 436-459.

Smetana, J. (1985). 'Preschool Children's Conceptions of Transgressions: Effects of Varying Moral and Conventional Domain-Related Attributes', *Developmental Psychology*, 21, 18-29.

Smith, M. (1987). 'The Humean Theory of Motivation', *Mind:New Series*, 96:381, 36-61.

Smith, M. (1994). *The Moral Problem*. Blackwell: Oxford UK and Cambridge USA.

Smith, M. (2004). *Ethics and the A Priori: Selected Essays on Moral Psychology and Meta-Ethics*. Cambridge: Cambridge University Press.

Sobel, D. (1999). 'Do the Desires of Rational Agents Converge?', *Analysis*, 59:3, 137-147.

Stocker, M. (1979). 'Desiring The Bad – An Essay In Moral Psychology', *The Journal Of Philosophy*, 738-753.

Stocker, M. with Hegeman, E. (1996). *Valuing Emotions*. Cambridge: Cambridge University Press.

Stoljar, D. (2005). 'Physicalism and Phenomenal Concepts', *Mind and Language*, 20:5, 469-494.

Sturgeon, S. (2007). 'Normative Judgement', *Philosophical Perspectives*, 21, 569-5877.

Taylor, S. E. (1989). *Positive Illusions: Creative Self-Deception and the Healthy Mind*. New York: Basic Books.

Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of Phenomenal Mind*. MIT Press.

Tye, M. (2009). *Consciousness Revisited: Materialism without Phenomenal Concepts*. Cambridge: MIT Press.

van Roojen, M. (2011). 'Moral Cognitivism vs. Non-Cognitivism', *The Stanford Encyclopedia of Philosophy (Spring 2011 Edition)*, Zalta, E. N. (ed.), URL = <http://plato.stanford.edu/archives/spr2011/entries/moral-cognitivism/>.

Velleman, J. D. (1992). 'The Guise of the Good', *Noûs*, 26:1, 3-26.

Vlastos, G. (1973). 'The Individual as an Object of Love in Plato', *Platonic Studies*. Princeton University Press, 1-34.

Watson, G. (1975). 'Free Agency', *Journal of Philosophy*, 72:8, 205-220.

Williams, B. (1973). *Utilitarianism: For and Against*. Cambridge: Cambridge University Press.

Williams, B. (1979). 'Internal and External Reasons' in his *Moral Luck* (1981). Cambridge: Cambridge University Press, 101-13.

Williams, B. (1985). *Ethics and the Limits of Philosophy*. Fontana Press.

Yim M. Y., Aertsen A., and Kumar A. (2011). 'Significance of Input Correlations in Striatal Function', *PLOS Computational Biology*, 7:11.

Young, L., Koenigs, M, Kruepke, and Newman, J. P. (2012). 'Psychopathy Increases Perceived Moral Permissibility of Accidents', *Journal of Abnormal Psychology*, 121:3, 659-667.

Zalla, T., Barlassina, L., Buon, M, and Leboyer, M. (2011). 'Moral Judgment in Adults with Autism Spectrum Disorders', *Cognition*, 121, 115–126.